

UC Irvine ISI-BUDS 2023 Day 09: GLM

Zhaoxia Yu

2023-07-20

Learning Objectives

Learning Objectives

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Review of LM
- ▶ GLM
 - ▶ Logistic Regression
 - ▶ Poisson Regression
 - ▶ Multinomial Regression
- ▶ The assumption of independent observations

Review of LM

A Linear Model (LM)

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

► Suppose

$$Y = \beta_0 + x_1 \times \beta_1 + \dots + x_p \times \beta_p + \epsilon,$$

where

- the regressand Y is the response / outcome / dependent / endogenous variable
- the regressors (x_1, \dots, x_p) are the p covariates / independent / explanatory variables
- the random term ϵ has a zero mean and variance $\sigma^2 > 0$
- the intercept is β_0 , the other p coefficients are β_1, \dots, β_p

A Linear Model (LM)

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Consider the i th observation:

$$Y_i = \beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p + \epsilon_i, i = 1, \dots, n$$

- ▶ Basic assumptions

- ▶ $E(\epsilon_i) = 0$, which is equivalent to

$$E(Y_i|X_i) = \beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p$$

- ▶ $\text{Var}(\epsilon_i) = \sigma^2$. Note, this is equivalent to say

$$\text{Var}(Y_i|X_i) = \sigma^2.$$

- ▶ $(\epsilon_1, \dots, \epsilon_n)$ are mutually independent

- ▶ If $(\epsilon_1, \dots, \epsilon_n)$ are i.i.d. $N(0, \sigma^2)$, we can derive t-tests and F-tests

- ▶ Question: what if the assumptions are violated?

Logistic Regression

A Motivating Example of GLM

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

- ▶ A motivating example: Consider a binary response variable, i.e., Y_i takes values of 0 or 1.
- ▶ Is LM a good choice for this problem?

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

A Motivating Example of GLM (continued)

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

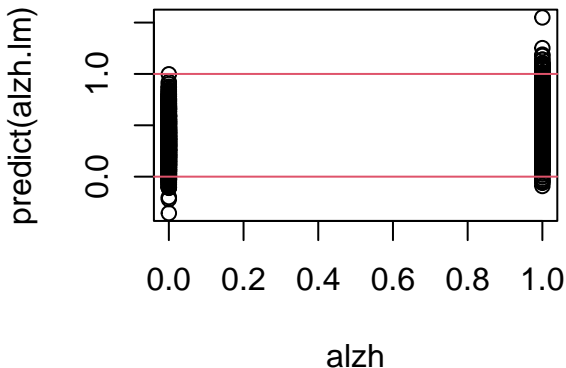
The Assumption
of Independence

- ▶ Consider the Alzheimer data
- ▶ We create a binary variable

```
alzheimer=read.csv("alzheimer_data.csv", header = TRUE)
#dim(alzheimer)
#names(alzheimer)
attach(alzheimer)
#length(unique(id))
alzh=(diagnosis>0)*1 #"*1" to create a 0-1 variable
```

A Motivating Example of GLM (continued)

```
alzh.lm = lm(alzh ~ age + female + educ+lhippo + rhippo)
par(mar = c(4, 4, 0.5, 0.5))
plot(alzh, predict(alzh.lm)); abline(h=c(0,1), col=2)
```



UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

A Motivating Example of GLM (continued)

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Is `alzh.lm` a good model for `alzh`?
- ▶ Several assumptions of the LM have been violated, and
- ▶ The predicted values using LM are not between 0 and 1!
- ▶ Let $X_i = (x_{i1}, \dots, x_{ip})^T$, i.e., the vector of covariates for the i th subject.
- ▶ Let $\pi_i = E(Y_i|X_i)$, the expected probability. We would like to make sure that $\pi_i \in [0, 1]$
- ▶ How?

Logistic Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Consider the a special transformation of π_i :

$$\text{logit}(\pi_i) = \log \frac{\pi_i}{1 - \pi_i} \in (-\infty, \infty)$$

- ▶ This is the so-called “logit” link!
- ▶ $\pi_i = E[Y_i|X_i]$: probability of having AD for a subject with covariates X_i .
- ▶ $\frac{\pi_i}{1-\pi_i}$: odds
- ▶ $\text{logit}(\pi_i) = \log \frac{\pi_i}{1-\pi_i} = \log \frac{P(Y_i=1|X_i)}{P(Y_i=0|X_i)}$: log-odds!

Logistic Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ We connect π_i and a linear function of the covariates X_i by assuming

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p$$

- ▶ Essentially, we model the log-odds.
- ▶ But Y_i is a random variable. We need a distribution. A natural choice is the Bernoulli distribution

$$Y_i | X_i \sim \text{Bernoulli}(\pi_i)$$

- ▶ pmf, mean, variance:

Logistic Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Estimation of is typically conducted by maximizing the corresponding likelihood function
- ▶ How to obtain the likelihood function
 - ▶ $E(Y_i|X_i) = \pi_i = \frac{\exp\{\beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p\}}{1 + \exp\{\beta_0 + x_{i1} \times \beta_1 + \dots + x_{ip} \times \beta_p\}}$
 - ▶ $f(Y_i|X_i) = \pi_i^{Y_i}(1 - \pi_i)^{1 - Y_i}$, i.e.,
 - ▶ $f(Y_i|X_i) = \pi_i$ if $Y_i = 1$
 - ▶ $f(Y_i|X_i) = 1 - \pi_i$ if $Y_i = 0$
 - ▶ independence: $f(Y|X) = \prod_{i=1}^n f(Y_i|X_i)$
 - ▶ $L(\beta_0, \beta_1, \dots, \beta_p) = f(Y|X)$

Logistic Regression for Retrospective Studies

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ In the previous slide we model $E(Y_i|X_i)$. Because Y_i is binary, we have $E(Y_i|X_i) = \Pr(Y_i = 1|X_i)$.
- ▶ Retrospective studies are often considered because a prospective study might take many years and is costly.
- ▶ In a retrospective study, subjects are recruited based on their disease status. Let $z = 1$ denote being sampled and $z = 0$ otherwise. Let

$$\Pr(z = 1|y = 0) = p_0$$

$$\Pr(z = 1|y = 1) = p_1$$

- ▶ For a retrospective study, the logistic regression models $\Pr(y = 1|z = 1, x)$

Logistic Regression for Retrospective Studies

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Does this affect the interpretation of the parameters?
- ▶ Let $\theta = Pr(y = 1|x)$ and $\phi = Pr(y = 1|z = 1, x)$. By Bayes' theorem and assuming that z does not depend on x ,

$$\begin{aligned}\phi &= Pr(y = 1|z = 1, x) \\&= \frac{Pr(z = 1|y = 1, x)Pr(y = 1|x)}{Pr(z = 1|y = 1, x)Pr(y = 1|x) + Pr(z = 1|y = 0, x)Pr(y = 0|x)} \\&= \frac{p_1\theta}{p_1\theta + p_0(1 - \theta)}\end{aligned}$$

Logistic Regression for Retrospective Studies

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- Therefore

$$\log\left(\frac{\phi}{1-\phi}\right) = \log\left(\frac{p_1\theta}{p_0(1-\theta)}\right) = \log(p_1/p_0) + \log\left(\frac{\theta}{1-\theta}\right)$$

- The result suggests that, when using logistic regression,
 - the only difference between a prospective study and a retrospective study would be the intercept.
 - the inference for the other parameters is still valid even though the subjects were recruited based on their disease status (such as a retrospective case-control study)

Logistic Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ How to obtain the maximum likelihood estimates (MLE) of the parameters $(\beta_0, \dots, \beta_p)$?
 - ▶ Iteratively re-weighted least squares (IRLS): the default method used by R
 - ▶ The Newton-Raphson algorithm

The Motivating Example of Logistic Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

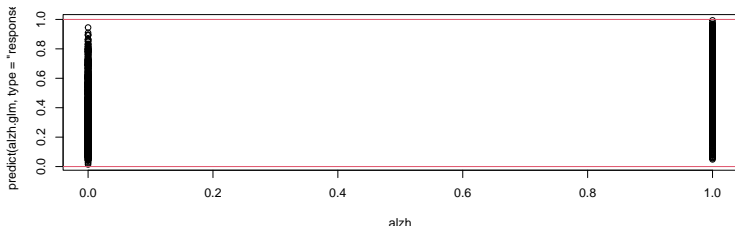
Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
alzh.glm = glm(alzh ~ age + female + educ+lhppo, family=binomial)
par(mar = c(4, 4, 0.5, 0.5))
plot(alzh, predict(alzh.glm, type="response")); abline(h=c(0,1), col=2)
```



#More visualizations

#<https://blogs.uoregon.edu/rclub/2016/04/05/plotting-your-logistic-regression-models/>

Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
summary(alzh.glm)$coefficients[-1,]
```

##	Estimate	Std. Error	z value	Pr(> z)
## age	0.01813761	0.004246088	4.271605	1.940715e-05
## female	-1.32020475	0.096534651	-13.675968	1.413151e-42
## educ	-0.05640342	0.013279326	-4.247461	2.162067e-05
## lhppo	-1.98502114	0.114028821	-17.408065	7.166544e-68

Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

► What is the β_0 in a logistic regression?



Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Consider the age variable. The estimated coefficient is 0.018138. What information does it provide?
- ▶ The estimated log-odds AD for subject i is (or add a constant **determined by study design**, as the previous slides)

$$\text{logit}(\hat{\pi}_i) = \hat{\beta}_0 + \hat{\beta}_{\text{age}} \text{age}_i + \hat{\beta}_2 \text{female}_i + \hat{\beta}_3 \text{educ}_i + \hat{\beta}_4 \text{hippo}_i$$

- ▶ Let $\tilde{\pi}_i$ denote estimated log-odds after one year

$$\text{logit}(\tilde{\pi}_i) = \hat{\beta}_0 + \hat{\beta}_{\text{age}}(\text{age}_i + 1) + \hat{\beta}_2 \text{female}_i + \hat{\beta}_3 \text{educ}_i + \hat{\beta}_4 \text{hippo}_i$$

Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ The estimated change in log-odds

$$\text{logit}(\tilde{\pi}_i) - \text{logit}(\hat{\pi}_i) = \log \frac{\tilde{\pi}_i}{1 - \tilde{\pi}_i} - \log \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} = 0.018138$$

- ▶ Take exponential of both sides, we have

$$\frac{\frac{\tilde{\pi}_i}{1 - \tilde{\pi}_i}}{\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}} = \exp(0.018138)$$

- ▶ The odds of AD in one year later is $\exp(0.018138) = 1.018303$ times of the current odds.
- ▶ The estimated increase in odds of AD in a year is $e^{0.018138} - 1 = 1.8303\%$

Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ A 95% confidence interval
 - ▶ First, obtain a 95% C.I. for the difference in log-odds:
 $(0.018138 - 1.96 * 0.004246, 0.018138 + 1.96 * 0.004246) = (0.00982, 0.0265)$
 - ▶ Then, we transform them to increase in odds:
 $(e^{0.00982} - 1, e^{0.0265} - 1) = (0.99\%, 2.69\%)$

Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ What if we are interested in the increase in odds of AD in ten years (everything else is fixed)?
- ▶ The estimated increase in odds of AD in 10 years is

$$e^{10 \cdot 0.018138} - 1 = 19.89\%$$

- ▶ A 95% C.I. for 10-year increase in odds:

```
exp(10*c(0.018138-1.96*0.004246, 0.018138+1.96*0.004246))-1
```

```
## [1] 0.1031375 0.3029118
```

i.e., (10.3%, 30.3%)

Interpreting a logistic regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Very often, we also want to know the significance of a variable after adjusting for other important covariates?
- ▶ Does age show a significant effect after adjusting for gender, education, and hippocampus volume?
- ▶ A test for $H_0 : \beta_{age} = 0$ using the Wald test (a type of large-sample test)

```
summary(alzh.glm)$coefficients["age",]
```

```
##      Estimate  Std. Error      z value    Pr(>|z|)
## 1.813761e-02  4.246088e-03  4.271605e+00  1.940715e-05
```

- ▶ Other tests, such as likelihood ratio test, can also be used

- ▶ Recall that we used the logit link in the logistic regression

$$g(\pi_i) = \text{logit}(\pi_i) = \frac{\pi_i}{1 - \pi_i},$$

where $\pi_i = E(Y_i|X_i)$.

- ▶ How about LM? $g(\mu_i) = \mu_i$, i.e., LM uses the identity link
- ▶ Poisson $g(\lambda_i) = \log(\lambda_i)$. $Y_i|X_i \sim \text{Poisson}(\lambda_i)$.

Poisson Regression

Poisson Regression: The Model

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Poisson regression is often used to model count data
- ▶ Why are count data special?
 - ▶ Count data are non-negative
 - ▶ Count data take integer values
- ▶ Count data often violate the assumption of “constant variance”
 - ▶ Count data often follow a Poisson distribution
 - ▶ Consider $K \sim \text{Poisson}(\lambda)$. $E(K) = ?$, $\text{Var}(K) = ?$, pmf ?

Poisson Regression: Motivating Example

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Neurons may fire selectively for particular types of stimuli
- ▶ To understand whether a neuron is a visual-selective neuron, 20 trials were run for each of the five image categories:
 - ▶ animal, fruit, kids, military, space
- ▶ In each trial, the number of spikes (the number of times that the neuron fired) within a 1-second window was recorded

Poisson Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
library(tidyverse)
#https://www.ics.uci.edu/~zhaoxia/Data/chosen_neuron_data.csv
chosen_neuron_data <- read_csv(
  "https://www.ics.uci.edu/~zhaoxia/Data/chosen_neuron_data.csv")
chosen_neuron_data <- chosen_neuron_data[, c(2:4)]
dim(chosen_neuron_data)
```

```
## [1] 100 3
```

```
names(chosen_neuron_data)
```

```
## [1] "trial_number" "n_spikes" "image_categ"
```

Poisson Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
attach(chosen_neuron_data)
# Even split of image categories among trials
table(image_categ)
```

```
## image_categ
##   Animal   Fruit   Kids Military   Space
##      20      20      20      20      20
```

```
sapply(split(n_spikes, image_categ), mean)
```

```
##   Animal   Fruit   Kids Military   Space
##   0.05    3.60    0.15    0.25    0.05
```


Poisson Regression: Visualize the count data (by image category)

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

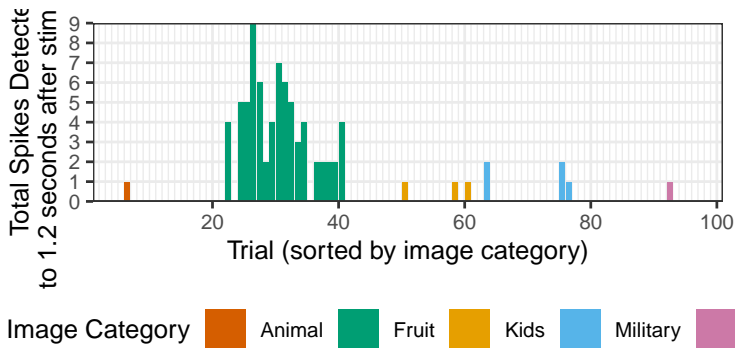
Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
neuron_trial_plot
```



Poisson Regression

Zhaoxia Yu

Learning Objectives

Review of LM

Logistic Regression

The Assumption of Independence

[illegible]

Poisson Regression: Model Summary

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
poisson_neuron_table
```

##	Estimate	Std. Error	z value	Pr(> z)
## Animal	-2.995732	0.9999998	-2.995733	2.737861e-03
## Fruit	1.280934	0.1178511	10.869084	1.618171e-27
## Kids	-1.897120	0.5773503	-3.285908	1.016541e-03
## Military	-1.386294	0.4472132	-3.099851	1.936181e-03
## Space	-2.995732	0.9999998	-2.995733	2.737861e-03

Poisson Regression: Visualize Observed v.s. Fitted

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

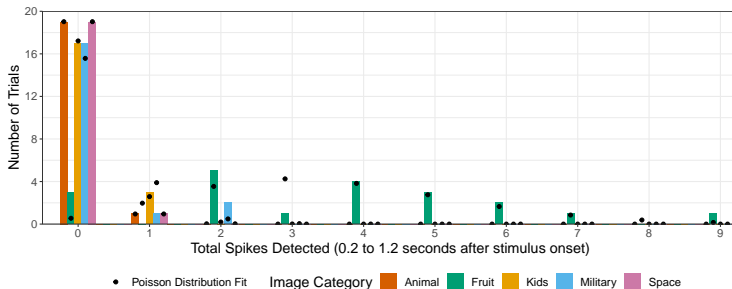
Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

poisson_obs_fit_plt



The Deviance of GLM object

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Next, we would like to discuss the significance of the `image_categ` variable. To do so, we first look at the deviance of a GLM object
- ▶ The deviance of a GLM object `obj` is

$$2[\log(L_{saturated}) - \log(L_{obj})]$$

- ▶ What is the null hypothesis of no visual-selection?

$$H_0 : \beta_{Animal} = \beta_{Fruit} = \beta_{Kids} = \beta_{Military} = \beta_{Space}$$

- ▶ What is the d.f. in a likelihood ratio test (LRT)?

Poisson Regression: The Overall Significance

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Consider two nested models obj1 and obj2 , the difference in their deviances is

$$2[\log(L_{\text{obj2}}) - \log(L_{\text{obj1}})],$$

which is the LRT statistic.

- ▶ What is the saturated model?
 - ▶ Logistic: $\pi_i = y_i$ and $L_{\text{saturated}} = 1$
 - ▶ Poisson: $\lambda_i = y_i$ and $L_{\text{saturated}} = \prod_i \frac{y_i^{y_i} e^{-y_i}}{y_i!}$.

Poisson Regression: The Overall Significance

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
# Test for visual selectivity: Likelihood Ratio Test
```

```
poisson_fit0 = glm(n_spikes ~ 1, data=chosen_neuron_data, family=poisson(link="log"))  
anova(poisson_fit0, poisson_fit, test = "LRT")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: n_spikes ~ 1
```

```
## Model 2: n_spikes ~ image_categ - 1
```

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1          99      260.985
```

```
## 2          95       81.213  4   179.77 < 2.2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Poisson Regression: The Overall Significance

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
# Test for visual selectivity: Rao's score test
anova(poisson_fit0, poisson_fit, test = "Rao")
```

```
## Analysis of Deviance Table
##
## Model 1: n_spikes ~ 1
## Model 2: n_spikes ~ image_categ - 1
##   Resid. Df Resid. Dev Df Deviance   Rao Pr(>Chi)
## 1         99    260.985
## 2         95    81.213  4   179.77 236.3 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Poisson Regression: The Overall Significance

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
# Test for visual selectivity: Wald test. I will explain reparameterization in a few slides
poisson_fit_repara <- glm(n_spikes ~ image_categ, data = chosen_neuron_data,
                          family = poisson(link="log"))
Wald.stat=poisson_fit_repara$coefficients[2:5] %*%
  solve (summary(poisson_fit_repara)$cov.unscaled[2:5,2:5]) %*%
  poisson_fit_repara$coefficients[2:5]
1-pchisq(Wald.stat, df=4)
```

```
##      [,1]
## [1,]    0
```

Poisson Regression: Model Interpretation

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

```
summary(poisson_fit)$coefficients
```

	Estimate	Std. Error	z value	Pr(> z)
## image_categAnimal	-2.995732	0.9999998	-2.995733	2.737861e-03
## image_categFruit	1.280934	0.1178511	10.869084	1.618171e-27
## image_categKids	-1.897120	0.5773503	-3.285908	1.016541e-03
## image_categMilitary	-1.386294	0.4472132	-3.099851	1.936181e-03
## image_categSpace	-2.995732	0.9999998	-2.995733	2.737861e-03

- ▶ $\hat{\beta}_{Fruit} = 1.2809$: What does it tell us?
- ▶ Recall that we used the log link

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

Poisson Regression: Model Interpretation

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Note that the model `poisson_fit` does not include β_0 .
- ▶ That's why we can estimate β_{Animal} , β_{Fruit} , β_{Kids} , $\beta_{Military}$, β_{space} .
- ▶ Question: how should we interpret the estimated coefficients if the intercept term was included?
 - ▶ Try `poisson_fit_repara <- glm(n_spikes ~ image_categ, data = chosen_neuron_data, family = poisson(link="log"))`
 - ▶ Are the two models equivalent? (Lab activity)

Poisson Regression: Model Interpretation

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Re-parameterization
- ▶ Parameters:
 - ▶ `poisson_fit`:
 - ▶ `poisson_fit_repara`:
- ▶ Compare the summary of the two models:
- ▶ Are they different models?

Parameterization 1: without intercept

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

```
summary(poisson_fit)
```

```
##
## Call:
## glm(formula = n_spikes ~ image_categ - 1, family = poisson(link = "log"),
##      data = chosen_neuron_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6833  -0.5876  -0.3162  -0.3162   2.3861
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## image_categAnimal    -2.9957     1.0000  -2.996  0.00274 **
## image_categFruit      1.2809     0.1179  10.869 < 2e-16 ***
## image_categKids      -1.8971     0.5774  -3.286  0.00102 **
## image_categMilitary  -1.3863     0.4472  -3.100  0.00194 **
## image_categSpace     -2.9957     1.0000  -2.996  0.00274 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 264.439  on 100  degrees of freedom
## Residual deviance:  81.213  on  95  degrees of freedom
## AIC: 163.18
##
## Number of Fisher Scoring iterations: 6
```

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

Parameterization 2: with intercept

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

```
summary(poisson_fit_repara)
```

```
##
## Call:
## glm(formula = n_spikes ~ image_categ, family = poisson(link = "log"),
##      data = chosen_neuron_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6833  -0.5876  -0.3162  -0.3162   2.3861
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.996e+00  1.000e+00  -2.996  0.00274 **
## image_categFruit  4.277e+00  1.007e+00   4.247  2.16e-05 ***
## image_categKids   1.099e+00  1.155e+00   0.951  0.34139
## image_categMilitary 1.609e+00  1.095e+00   1.469  0.14178
## image_categSpace  -3.140e-16  1.414e+00   0.000  1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 260.985  on 99  degrees of freedom
## Residual deviance:  81.213  on 95  degrees of freedom
## AIC: 163.18
##
## Number of Fisher Scoring iterations: 6
```

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

Linear Functions of Parameters

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ The Poisson regression we fit provides estimates of β_{Animal} , β_{Fruit} , β_{Kids} , $\beta_{Military}$, β_{space} , which are the log of the Poisson rates
- ▶ What if we are interested in difference between specific groups? e.g.,
 - ▶ $\beta_{Fruit} - \beta_{Animal}$
 - ▶ $\frac{\beta_{Fruit} + \beta_{Animal} + \beta_{Kids} + \beta_{Military} + \beta_{Space}}{5}$
 - ▶ $\beta_{Fruit} - \frac{\beta_{Animal} + \beta_{Kids} + \beta_{Military} + \beta_{Space}}{4}$
- ▶ They are linear functions of the coefficients, i.e., in the form of $a^T \beta$, where a is a 5-by-1 vector.

Linear Functions of Parameters

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ LM/GLM provides not only estimated coefficients but also the variance-covariance of the estimated covariates
 - ▶ Let $\hat{\beta} = c(\hat{\beta}_1, \dots, \hat{\beta}_p)^T$
 - ▶ Let $\hat{\Sigma}$ denote the estimated variance-covariance of $\hat{\beta}$
 - ▶ Let a be linear coefficients

Inference of Linear Functions of Parameters

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Consider a linear function : $a^T \beta$
- ▶ Estimate: $a^T \hat{\beta}$
- ▶ Variance of the estimate: $Var(a^T \hat{\beta}) = a^T \hat{\Sigma} a$
- ▶ Standard Error (SE): $s.e.(a^T \hat{\beta}) = \sqrt{a^T \hat{\Sigma} a}$
- ▶ A 95% confidence interval:

$$(a^T \hat{\beta} - 1.96 * s.e.(a^T \hat{\beta}), a^T \hat{\beta} + 1.96 * s.e.(a^T \hat{\beta}))$$

- ▶ Z-value: $\frac{a^T \hat{\beta} - a^T \beta}{s.e.(a^T \hat{\beta})}$

Inference of Linear Functions of Parameters: Example

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

► Parameter of interest: $\frac{\beta_{Fruit} + \beta_{Animal} + \beta_{Kids} + \beta_{Military} + \beta_{Space}}{5}$

```
a=matrix(rep(1/5,5), 1)
a%*%poisson_fit$coefficients #estimate
```

```
##           [,1]
## [1,] -1.598789
```

```
sqrt(a%*%summary(poisson_fit)$cov.unscaled%*%t(a)) #s.e.
```

```
##           [,1]
## [1,] 0.3192003
```

Linear Contrasts

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ Linear contrasts are a special family of linear functions
- ▶ We say $a^T \beta = \sum_i a_i \beta_i$ is a linear contrast if $\sum a_i = 0$, where $a = (a_1, \dots, a_p)^T$.
- ▶ Often, we are interested in whether a linear contrast is zero, i.e., $H_0 : a^T \beta = 0$
- ▶ z-value: $\frac{a^T \hat{\beta} - 0}{s.e.(a^T \hat{\beta})}$

Linear Contrast (e.g., Fruit vs Animal)

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
a <- matrix(c(-1, 1, 0, 0, 0), 1)
#estimate
fruit_animal_est = a%*%poisson_fit$coefficients
#variance
fruit_animal_var = a%*%summary(poisson_fit)$cov.unscaled%*%t(a)
#z value
print(fruit_animal_est/sqrt(fruit_animal_var))
```

```
##           [,1]
## [1,] 4.247274
```

Linear Contrast (e.g., Fruit vs Animal) with the multcomp package

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
library(multcomp)
a <- matrix(c(-1, 1, 0, 0, 0), 1)
t <- glht(poisson_fit, linfct = a)
summary(t)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Fit: glm(formula = n_spikes ~ image_categ - 1, family = poisson(link = "log"),
## data = chosen_neuron_data)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## 1 == 0 4.277 1.007 4.247 2.16e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Multinomial Logistic Regression

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

```
library(nnet)
multinom(diagnosis ~ age + female + educ + lhippo + rhippo)

## # weights:  21 (12 variable)
## initial value 2966.253179
## iter  10 value 2372.326777
## final value 2288.461323
## converged

## Call:
## multinom(formula = diagnosis ~ age + female + educ + lhippo +
##           rhippo)
##
## Coefficients:
## (Intercept)      age      female      educ      lhippo      rhippo
## 1      3.671844 0.026773068 -1.237127 -0.04694669 -1.251316 -0.4056775
## 2      8.147569 0.005473649 -1.473799 -0.06379482 -1.794730 -0.7967240
##
## Residual Deviance: 4576.923
## AIC: 4600.923
```

Other concerns

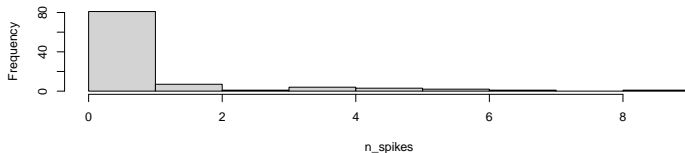
UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

- ▶ Dispersion: under- or over-dispersion
- ▶ Zero-inflated Poisson Regression
- ▶ Model selection ...

```
hist(n_spikes)
```

Histogram of n_spikes



*#Interested in how to fit a zero-inflated Poisson regression? See the link
#<https://www.rdocumentation.org/packages/pscl/versions/1.5.5/topics/zeroinfl>*

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

The Assumption of Independence

Independent Observations

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ The common assumption we have made in LM and GLM is that the observations are independent with each other
- ▶ This is not always the case
- ▶ Examples:

Independent Observations

UC Irvine
ISI-BUDS 2023
Day 09: GLM

Zhaoxia Yu

Learning
Objectives

Review of LM

Logistic
Regression

Poisson
Regression

The Assumption
of Independence

- ▶ What is the consequence of ignoring data independence?
 - ▶ The damage is probably worse than violations of distributions
 - ▶ Fortunately, tools have been developed to account for data dependence
 - ▶ Linear Mixed-Effects Model (LME)
 - ▶ Generalized Linear Mixed-Effects Model (GLMM)