

Canadian Bioinformatics Workshops

www.bioinformatics.ca
bioinformaticsdotca.github.io

Creative Commons

This page is available in the following languages:

Afrikaans Afrikaans Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macdonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik срpski (latinica) Sotho svenska
中文 華語 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French

[Learn how to distribute your work using this licence](#)

Generalized Linear Models



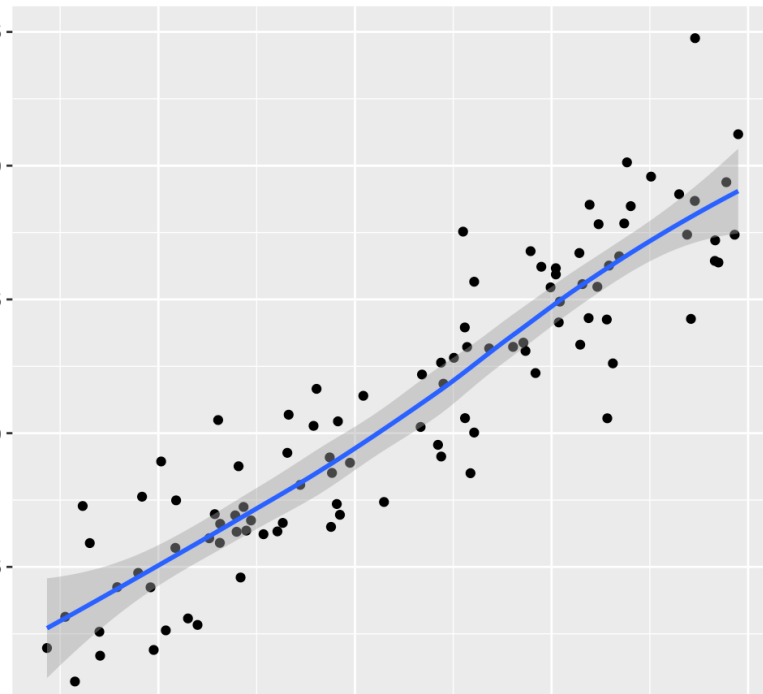
Some material adapted from 2021 lecture by Lauren Erdman

Learning Objectives

- By the end of this lecture, you will:
 - Learn to read tables into R and merge tables
 - Create a variety of publication-quality plots using `ggplot()` and use them to explore data
 - Learn how to fit a binary response variable using `glm()`

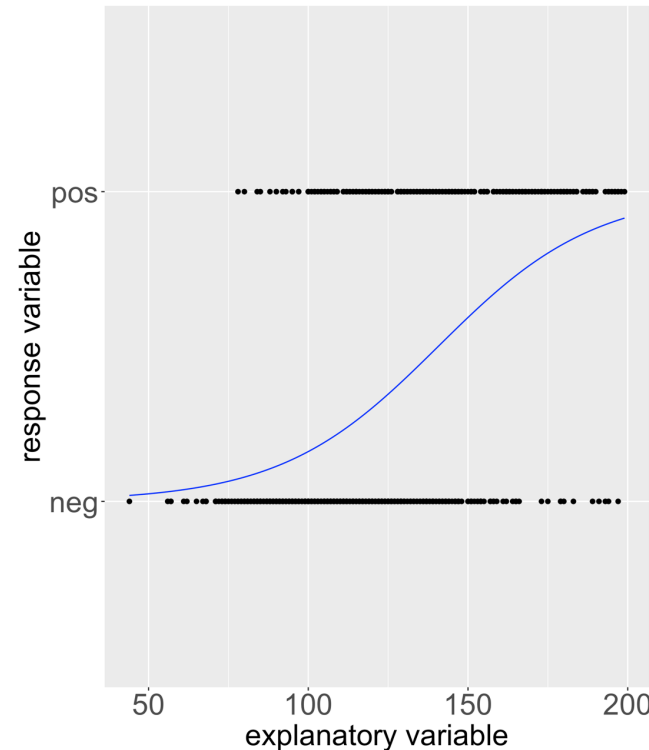
Generalized linear models (GLM) for binary response & count data

Continuous-valued x & y:
linear regression



`lm(y ~ x, data = df)`

Binary y, continuous-valued x:
logistic regression



`glm(y ~ x, data = df, family = "binomial")`

RNA-seq count data:
fit negative binomial
model

Important:
Model selection
must be data driven

Model fitting notation in R

$$Income = \beta_0 + \beta_1(PctLiteracy) + \epsilon$$

lm(Income ~ PctLiteracy,
data = dat)

$$diabetes = \beta_0 + \beta_1(glucose) + \beta_2(pregnancy) + \epsilon$$

glm(diabetes ~ glucose + pregnancy,
data = dat,
family = “**binomial**”)

Follow along with the worked example for
Module 3 ...

We are on a Coffee Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



HPC4Health



GenomeCanada