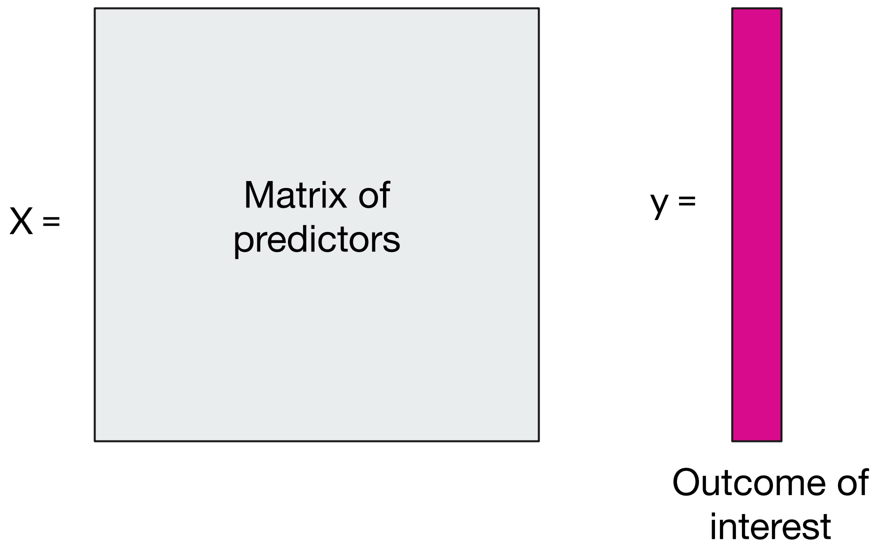
Chapter 12: Generalized Linear Models

Modern Clinical Data Science
Chapter Guides
Bethany Percha, Instructor

How to Use this Guide

- Read the corresponding notes chapter first
- Try to answer the discussion questions on your own
- Listen to the chapter guide (should be 15 min, max) while following along in the notes

Supervised Learning



Question 12.1

We saw the details of linear and logistic regression models in Chapters 8 and 9 and discussed the limitations of predictors' entering as a linear combination. What are some of those limitations?

$$\beta^T x = \beta_0 + \sum_{j=1}^{p} \beta_j x_j.$$

Question 12.2

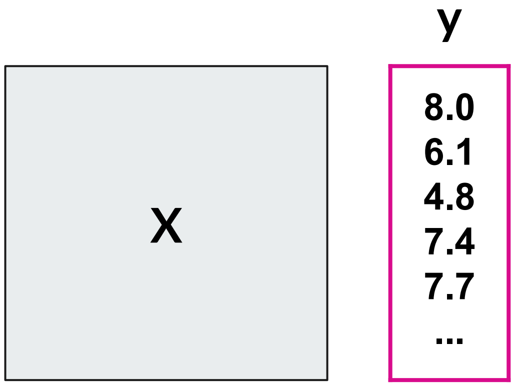
Just to confirm that you understand this notation, write out the form of $\beta^T x$ for a model with (a) one predictor, (b) three predictors. Write both the general form and the form for one training example, $x^{(i)}$.

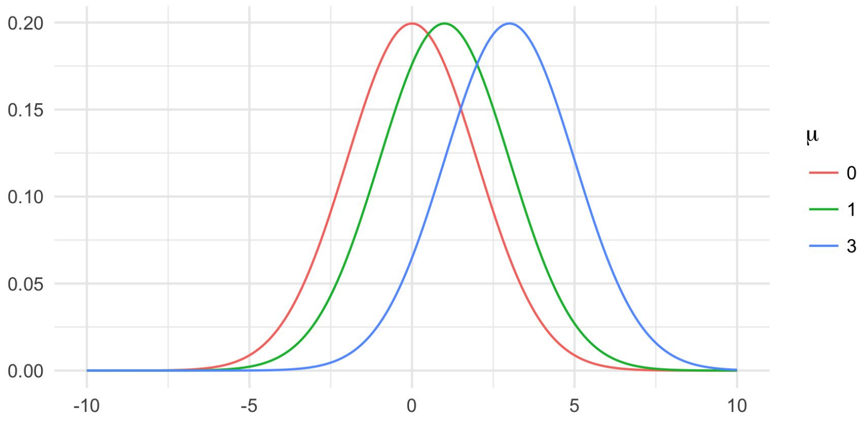
$$\beta^T x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

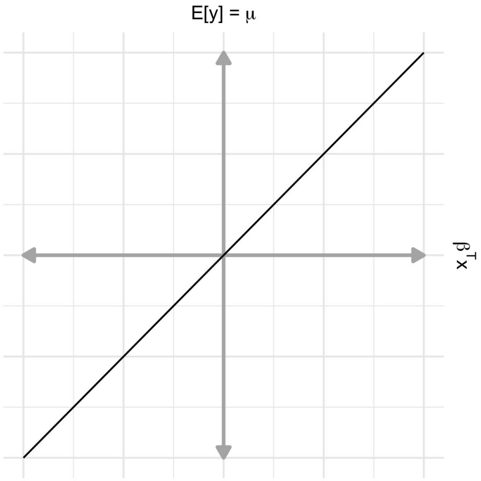
$$\beta^T x^{(i)} = \beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \beta_3 x_3^{(i)}$$

Linear Regression

$$E[y] = \mu = \beta^T x$$

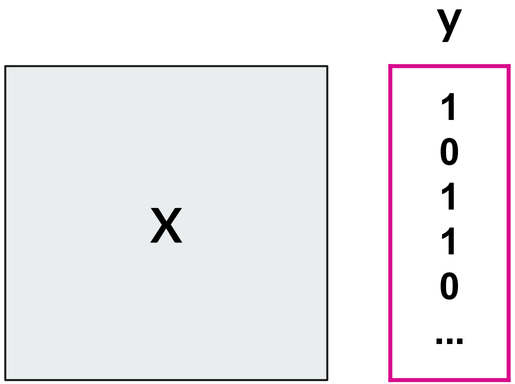


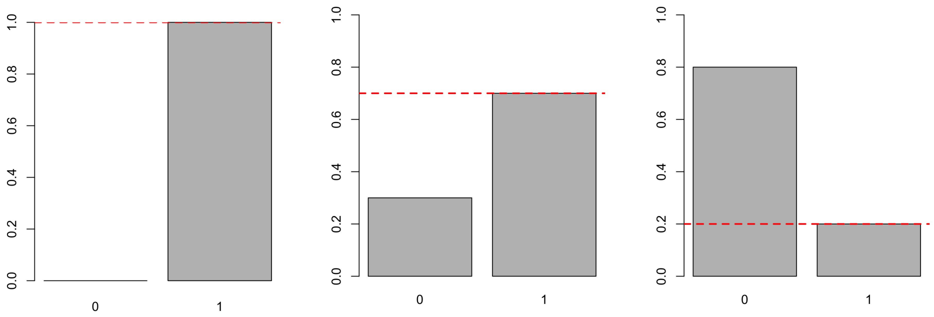


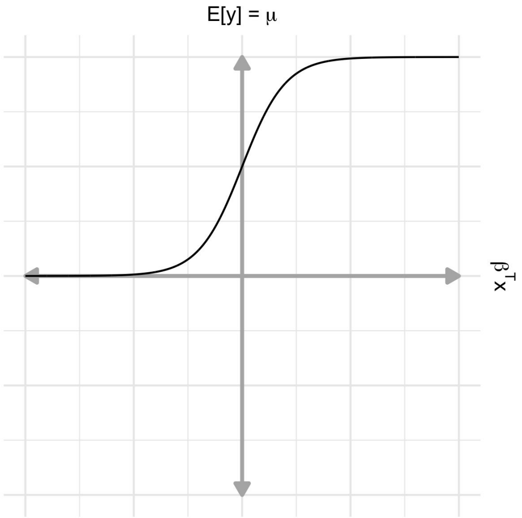


Logistic Regression

$$E[y] = \mu = \frac{1}{1 + \exp\left(-\beta^T x\right)}$$

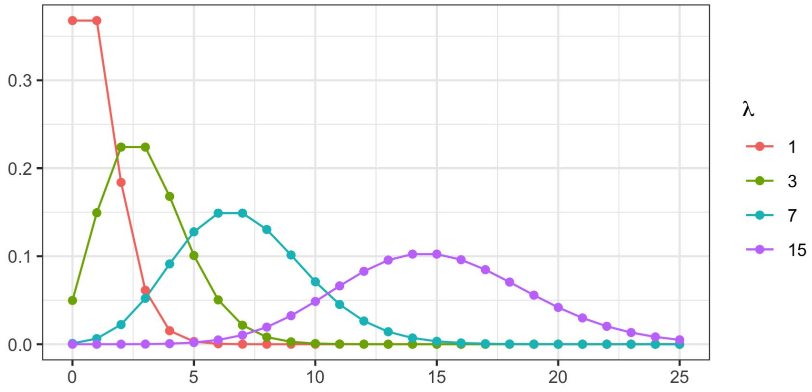


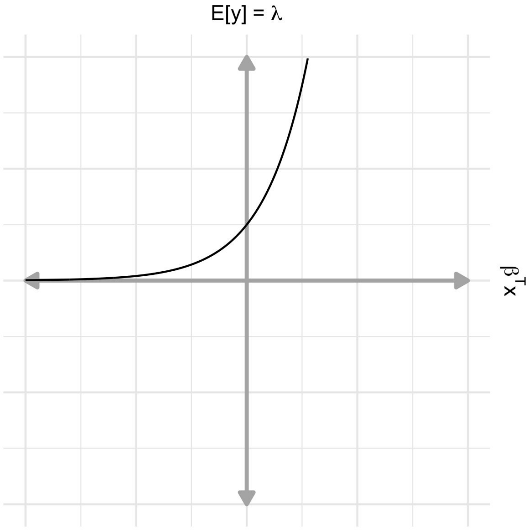


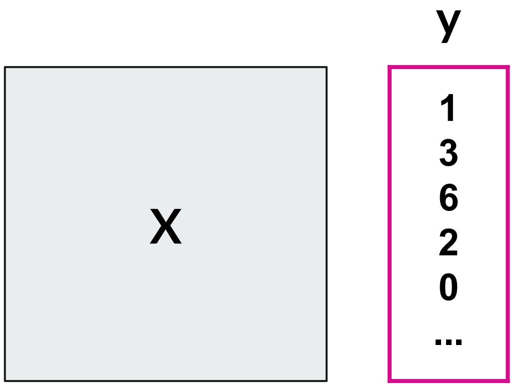


Poisson Regression

$$E[y] = \lambda = \exp(\beta^T x)$$







Question 12.3

In this model, how much does the mean of the outcome distribution, μ, change as you vary each predictor? For example, if you have $p = 3$ predictors, by how much does μ change as the value of x_2 changes by one unit (for example, from 1 to 2)? How much does μ change as the value of x_2 changes from 3 to 4? What about x_1 and x_3?

Question 12.4

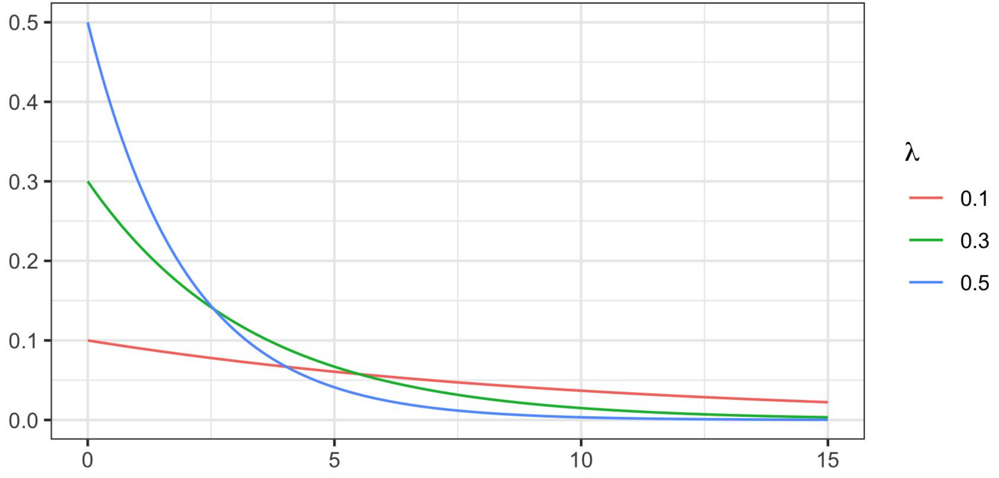
Let's revisit Question 9.1. Now that we've described logistic regression in the framework of GLMs, what more can you say about why the model is not of the form

$$\mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p?$$

Question 12.5

There are many other generalized linear models. In each case, the mean (expected value) of a probability distribution is related, via a link function, to a linear combination of the predictors.

Knowing this, how would you create a GLM where the outcome follows an exponential distribution (Section 4.7)? Which link would you use?

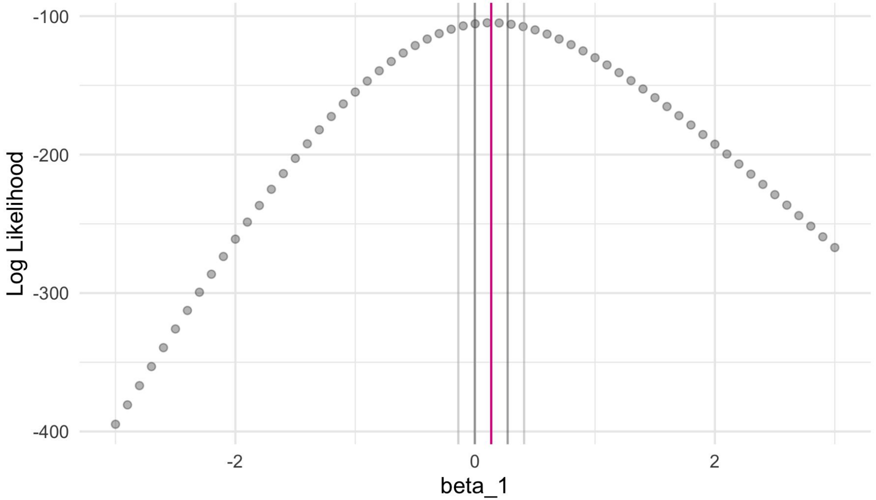


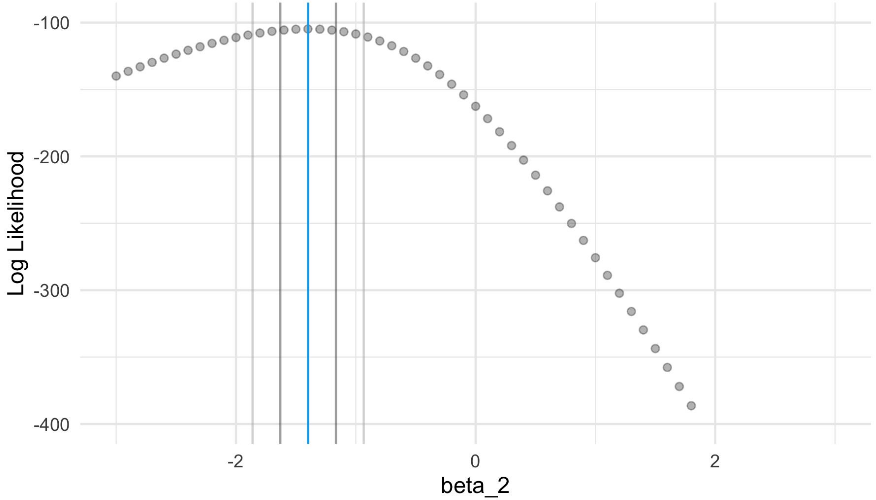
Question 12.7

Think of the log-likelihood as measuring the height of a hill. Your data,

$$\left\{x^{(1)},\ldots,x^{(n)}\right\}$$

don't change, so we don't care about their effect on the height. What we care about are the parameters, $\beta_0,\ldots,\beta_p$. For each combination of those $p+1$ parameters, the height changes. We want to find the combination of parameters





Question 12.8

There are many different numerical optimization algorithms that one can use to maximize the likelihood (i.e., find the top of the hill). One of them is called **Fisher scoring**. Examine the output of the logistic regression models in Chapter 9 and the Poisson regression model shown below in Section 12.6. Where do you see the term "Fisher scoring"? What do you think the term "Fisher scoring iterations" refers to?

```
Call:
glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.88232  -0.90614  -0.05965   0.86579   2.28489

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9780     0.2945   3.321 0.000897 ***
x1            0.1344     0.1372   0.980 0.327272
x2           -1.3981     0.2316  -6.035 1.59e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 277.26  on 199  degrees of freedom
Residual deviance: 209.54  on 197  degrees of freedom
AIC: 215.54

Number of Fisher Scoring iterations: 4
```

Question 12.9

These findings are reflected in the relative values of the Z-statistic (`z value`) and P-value (`Pr(>z|)|`) in the model output for the two coefficients. With that in mind, let's reconsider Question 9.6. How do these likelihood plots and the null distributions shown in Question 9.6 convey the same information?