**Chapter 2 – Analyzing a binary response – Part 2**

The previous chapter discussed:

* Estimating and making inferences about a probability of success
* Estimating and making inferences about a probability of success dependent on a two-level group variable

We now generalize this to a situation with a probability of success dependent on variable(s) with potentially many different possible levels. Again, estimation and inference will be our goal. We make these generalizations with the use of binary regression models.

**Section 2.1 – Linear regression models**

Review of normal linear regression models

Suppose Y is the response variable (a.k.a., dependent variable) and x1, …, xp are explanatory variables (a.k.a., independent variables or covariates). We relate Y to x1, …, xp through a normal linear regression model:

Y = β0 + β1x1 + β2x2 + … + βpxp + ε

where

* ε has a normal distribution with mean 0 and variance σ2; often referred to as the “error term”
* β0, …, βp are regression parameters

When we want to specify the ith value in a random sample, we can write this as

Yi = β0 + β1xi1 + β2xi2 + … + βpxip + εi

where i = 1, …, n and each εi is independent and has a normal distribution with mean 0 and variance σ2.

E(Y) is what one would expect Y to be on average in the population for a set of x1, x2, …, xp values:

E(Y) = β0 + β1x1 + β2x2 + … + βpxp.

Also, the model implies that each Y has a normal distribution with mean E(Y) and variance σ2.

Example: College and high school GPA

Y represents college GPA

x1 represents high school GPA

Model: E(Y) = β0 + β1x1

Other potential explanatory variables could be ACT score, involvement in high school activities, … .

An important aspect of these regression models is to interpret the effect an explanatory variable has on E(Y). Suppose we are interested in a specific explanatory variable xr.

* If βr is equal to 0, this says there is no linear relationship between the corresponding explanatory variable xr and E(Y).
* If βr > 0, there is a positive relationship.
* If βr < 0, there is a negative relationship.

The standard interpretation for the effect that xr has on E(Y) is

For every one-unit increase in xr, E(Y) increases by βr units when holding the other explanatory variables constant.

When a sample of size n is taken, the regression parameters are estimated by using the least squares method. This actually is equivalent to using maximum likelihood estimation when the likelihood function is the product of n different normal distributions.

The estimated normal linear regression model is



where , …,  are the estimated regression parameters. Note that “” is the standard notation used on the left side, but one could use something like .

The standard interpretation for the estimated effect that xr has on E(Y) is

For every one-unit increase in xr, E(Y) is estimated to increase by  units when holding the other explanatory variables constant.

Regression models for a binary response

Let Yi be a binary response variable for a random sample with Yi = 1 denotes a success, Yi = 0 denotes a failure, i = 1, …, n. For example, Yi could denote whether one graduates from college (1 for yes, 0 for no) and xi1 could denote the number of hours one studies per week.

Suppose Yi has a Bernoulli distribution, but now with the probability of success parameter E(Yi) = πi. The probability of success can be different for i = 1, …, n. Potentially, there could be n different parameters then we need to estimate!

We can simplify the number of parameters that we need to estimate by using a linear model of the form

 E(Yi) = β0 + β1xi1

where I am using just one explanatory variable to simplify the explanation. Because E(Yi) is πi, we could also write the model as

πi = β0 + β1xi1

Therefore, instead of potentially having n different parameters to estimate, we now only have two!!!

To estimate the regression parameters, we should not proceed as we would with normal linear regression models for the following reasons:

* Yi is binary here, but Yi had a continuous distribution in normal linear regression.
* Var(Yi) = πi(1 – πi) for a Bernoulli random variable; thus, the variance potentially changes for each Yi. With normal linear regression, Var(Yi) = Var(εi) = σ2 is constant for i = 1, …, n.

We could estimate the regression parameters through using maximum likelihood estimation that takes into account the Bernoulli distribution for Yi. The likelihood function is



where πi = β0 + β1xi1. Maximizing the likelihood function leads to maximum likelihood estimates of β0 and β1.

Unfortunately, there is still a problem: πi = β0 + β1xi1 is not constrained to be within 0 and 1. For particular values of β0, β1, and xi1, πi may end be greater than 1 or less than 0.

**Section 2.2 – Logistic regression models**

We can use a mathematical transformation in the model to prevent πi from being greater than 1 or less than 0. The most common transformation results in a specific type of binary regression model known as a logistic regression model:



Notice that  so that the numerator is always less than the denominator. Thus, 0 < πi <1.

The logistic regression model can also be written as



through using algebra (apply exp() to both sides of the above equation and then solve for πi). Notice that the left-hand side is the natural log transformation applied to the odds of a success! This will be very important for us later when interpreting the effect an explanatory variable has on the response variable.

Comments:

* + - The  part of the model is known as the linear predictor.
		- The  transformation is known as the logit transformation.
		- The most compact way to write the model is



* + - We can write the model without the i subscript when we want to state the model in general:







Obviously, this leads to some notational ambiguity with what we had in Section 1.1 for π, but the meaning should be obvious within the context of the problem.

Example: Plot of π vs. x (PiPlot.R)

When there is only one explanatory variable x1, β0 = 1, and β1 = 0.5 (or -0.5), a plot of π vs. x1 looks like the following:



We can make the following generalizations:

* 0 < π < 1
* When β1 > 0, there is a positive relationship between x1 and π.
* When β1 < 0, there is a negative relationship between x1 and π.
* The shape of the curve is somewhat similar to the letter s. This is referred to as a sigmoidal shape.
* The shape of the curve is exactly the same if it is rotated 180 degrees about π = 0.5; thus, the shape is “rotationally symmetric”.
* The slope of the curve is dependent on the value of x1. We can show this mathematically by taking the derivative with respect to x1: 

R code:

> par(mfrow = c(1,2))

> beta0 <- 1

> beta1 <- 0.5

> curve(expr = exp(beta0+beta1\*x)/(1+exp(beta0+beta1\*x)),

 xlim = c(-15, 15), col = "black", main = expression(beta[1] == 0.5), xlab = expression(x[1]), ylab = expression(pi))

> beta0 <- 1

> beta1 <- -0.5

> curve(expr = exp(beta0+beta1\*x)/(1+exp(beta0+beta1\*x)),

 xlim = c(-15, 15), col = "black", main = expression(beta[1] == -0.5), xlab = expression(x[1]), ylab = expression(pi))

Questions:

* What happens to the β1 = 0.5 plot when β1 is increased?
* What happens to the β1 = 0.5 plot when β1 is decreased to be close to 0?
* Suppose a plot of logit(π) vs. x1 was made. What would the plot look like?