Advanced STATS and R
GLM's
DAGs and DOODLES

The University of Oklahoma

The Exponential Family

- We have been looking at the linear model of the form:

$$E(Y_i) = \mu_i = x_i'\beta$$

Or $E(Y) = X\beta$ where X is the design matrix and x_i' is the ith row of the design matrix X

Where

$$Y_i \overset{\text{iid}}{\sim} N(\mu_i, \sigma^2)$$

This would be our typical multiple regression model.

Modification

- We can modify the form of this model to make it a little more general.
- Distribution no longer simply Normal but Exponential family
- $g(\mu_i) = x_i'\beta$
- g is called the link function

Exponential Family

- Suppose that the distribution of Y depends on only one parameter θ
- Then we could define Y to be distributed within the exponential family of distributions if:

$$f(y|\theta) = s(y)t(\theta)e^{a(y)b(\theta)}$$

- Or equivalently

$$f(y|\theta) = \exp\Big(\big(a(y)b(\theta) + c(\theta) + d(y)\big)\Big)$$

Another form of the exponential family called the exponential dispersion family (EDF).

$$f(y|\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

θ is called the canonical parameter – represents location

ϕ is the dispersion parameter and represents the scale

This will have some consequences

Mean and Variance for EDF

$$E(Y) = \mu = b'(\theta)$$
$$V(Y) = b''(\theta)a(\phi)$$

$b''(\theta)$ describes how the variance of the response relates to the mean.

In the case of the normal $b''(\theta) = 1$ and hence the variance is independent of the mean.

We shall do some more algebra later on the EDF

Modeling the GLM - 3 parts

The three components:

1. **A random response** Y that is a member of the exponential family (usually EDF)
2. **A linear predictor** $\eta = X\beta$
3. **A link function**

$$g(\mu_i) = \sum_j \beta_j x_{ij}\,, i = 1, \ldots, n$$

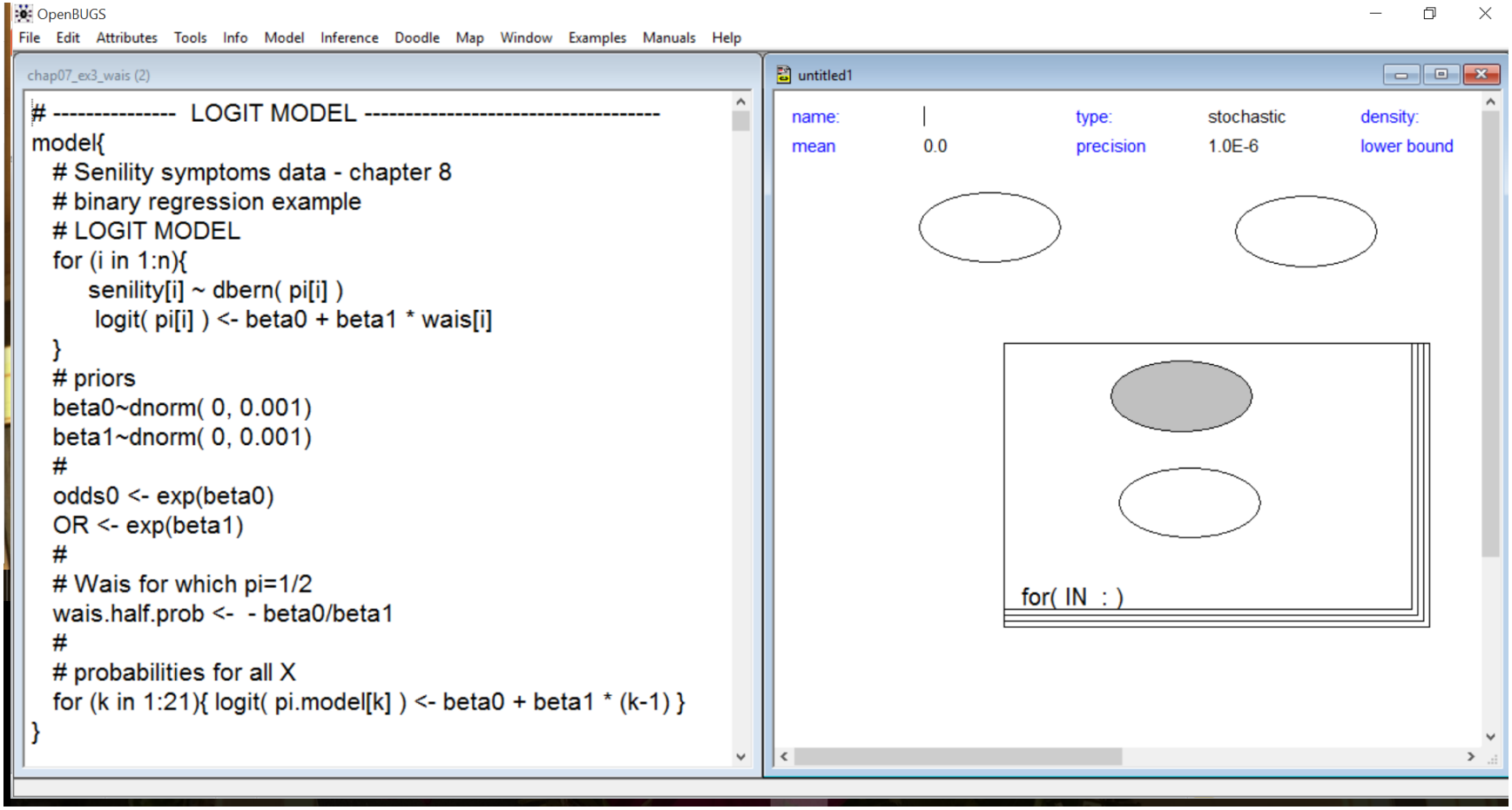
$$\mu_i = E(Y_i)$$

$$g(\mu_i) = \eta_i$$

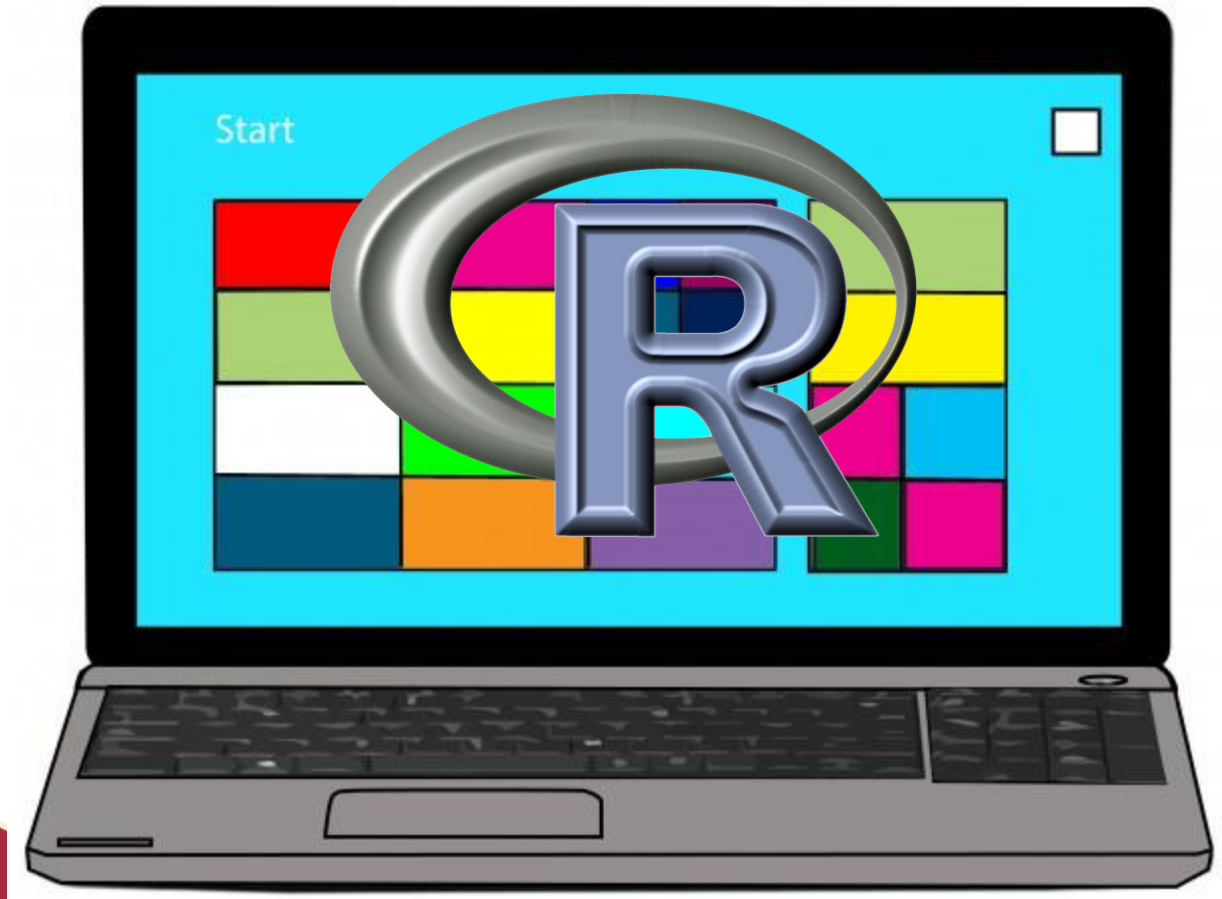
Canonical link

$$\eta = g(\mu) = \theta$$

Making a doodle from code



Bring your laptop or use Network.



Get the Data and Book

- Link below:
- statsandr.oucreate.com

Courses

- **Bayesian Stats MATH 4803/5803**
- Advanced Applied STATS MATH 4793/5793 (Next Year)
- **Applied Statistical Methods MATH 4753**