ADVANCED BAYESIAN MODELING

More About Variables

Transformations

A scalar transformation, such as $\log$, square, or square root, may be applied to an individual variable (response or explanatory).

A nonlinear scalar transformation of the response variable may make normal-theory assumptions more tenable.

New explanatory variables can be created by applying (nonlinear) transformations to the original ones (e.g., *polynomial regression*).

Centering and Standardization

An explanatory variable can be **centered** by subtracting its average:

$$x_i^{\text{cent}} = x_i - \bar{x}$$

(Do this only if there is an intercept, and it doesn't involve the variable.)
Sometimes improves Gibbs sampler convergence in Bayesian models (later).

The variable can further be **standardized** by dividing by its sample standard deviation:

$$x_i^{\text{stand}} = \frac{x_i - \bar{x}}{s_x}$$

Then its β coefficient will have the same units as the response.

Categorical Variables

A categorical (discrete, non-numeric) variable x with finitely many categories can be used as an explanatory variable:

For each category ℓ, create **indicator variable**

$$I_\ell(x_i) = \begin{cases} 1 & x_i \text{ is category } \ell \\ 0 & \text{otherwise} \end{cases}$$

Use these as explanatory variables (in X), excluding one of them if the regression has an intercept term.

(Alternatively, if categories of x are ordered, consider using a numeric coding.)

Interaction

A pure **interaction** between numerical explanatory variables x_1 and x_2 is the new variable whose values are

$$x_{i1}x_{i2} \qquad\qquad i = 1, \ldots, n$$

For computational reasons, it may help to center or standardize variables before forming their interaction.

Indicator variables of a categorical variable may participate in interactions.