Logistic Regression

When the dependent variable is a dummy variable

The logit or logistic model

- P is the probability of dummy = 1.
- $Ln(P) – Ln(1-P) = \alpha+\beta X$
- Behaves like exponential growth function near P=0.
  - Because Ln(1-near 0) is near 0.
- Behaves like upside-down exponential growth function near P=1.
  - Because Ln(near 1) is near 0.

The logit or logistic model

- P is the probability of dummy = 1.
- $Ln(P) - Ln(1-P) = \alpha+\beta X$
- $Ln(P/(1-P)) = \alpha+\beta X$

- $P = 1 / (1 + e^{-(\alpha+\beta X)})$

Logistic regression

- You can't use LS on $Ln(P) - Ln(1-P) = \alpha+\beta X$
- Because your data are all 0's and 1's. For either, you'd be taking the log of 0, which is undefined.
- Instead must use maximum likelihood
  - A computationally intense trial-and-error method
- Or group the data so that no group has 0.

Interpretation of β is not straightforward

- $Ln(P/(1-P)) = \alpha+\beta X$ is not linear in P
- So you can't say that when X changes by 1, the probability of being in the group increases by some amount or percentage.
- Instead, evaluate $P = 1 / (1 + e^{-(\alpha+\beta X)})$ at one X value and then at another and report what the two probabilities are.

Interpretation of β is not straightforward

- Instead, evaluate $P = 1 / (1 + e^{-(\alpha+\beta X)})$ at one X value and then at another and report what the two probabilities are.
- β as "odds ratio."
  - P/(1-P) is the "odds" of being in the group. E.g. The odds of heads on a coin toss are 1:1.
  - When X goes up by 1, the odds go up by about $\beta \times 100\%$. If before the odds were 1:1, the new odds would be about $(1+\beta)$:1.