PoliSci 4782 Political Analysis II

Binary Outcome Models

Seamus Wagner

The Ohio State University

Objectives

- Learn two types of binary outcome models (using different link functions):
  - logistic regression (logit)
  - probit regression
- Interpret their results
  - Model interpretation becomes more complicated in generalized linear models (depending on the actual link/mean function)

The Explosion of the Challenger Space Shuttle in 1986

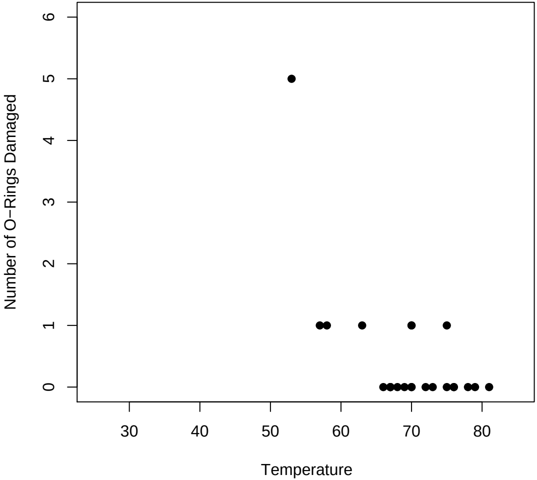
Background

- Space shuttle Challenger explodes 73 seconds after liftoff.
- All seven astronauts on board were killed.
- NASA had estimated probability of a shuttle accident at 1 in 100,000
- Investigation focuses on O-ring seals in the rocket boosters.
- At low temperatures, rubber gets brittle and less effective as a sealant.
- The launch was on a cold day (31° F).
- Could the disaster have been predicted? *Prevented?*

Let's Do Research

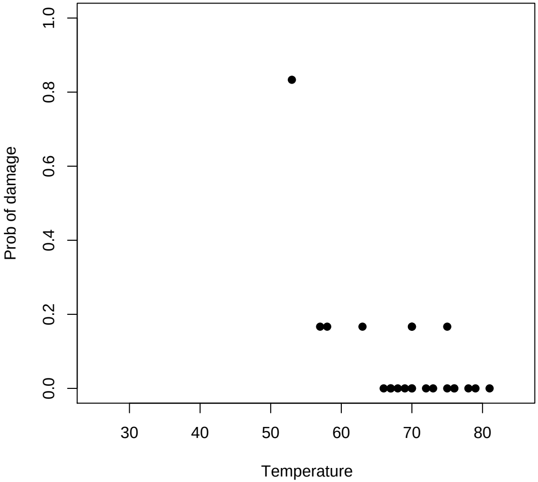
- NASA had data on O-ring damage and launch temperature from the 23 previous launches.
- 6 O-rings per shuttle.
- How better statistics could have saved 7 lives: *estimate probability of failure given temperature at launch was 31° F*:
  - Outcome variable: the number of O-rings (out of six) showing some damage
  - Explanatory variable: temperature at launch

Raw Data



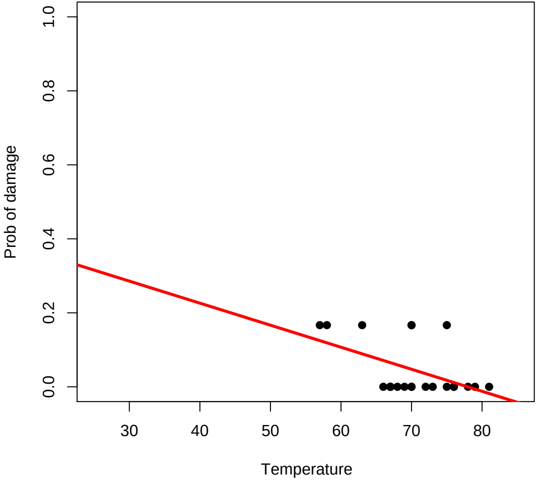
Number of O−Rings Damaged
0
1
2
3
4
5
6
30
40
50
60
70
80
Temperature

Probability for Catastrophic Failures



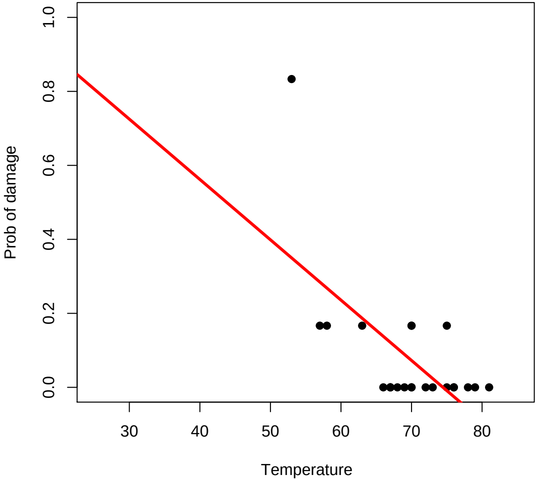
Just divide the outcome variable by 6.

One Mistake in Practice



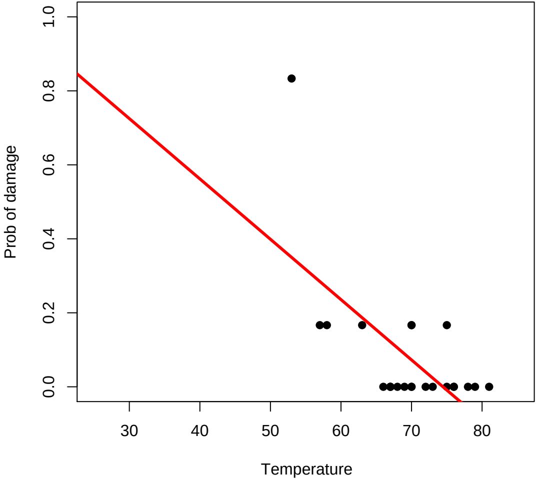
Many might drop the outlier *(by science never just do that without a strong reason!)*

How about Linear Regression?



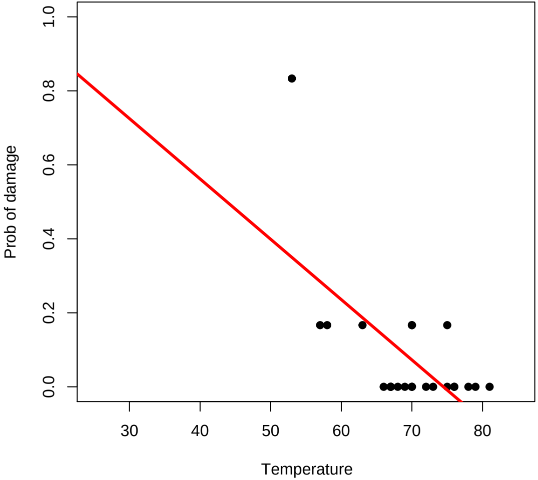
Problem: the linear model can predict $1 < Pr()$ and $0 > Pr()$.

Heteroskedasticity



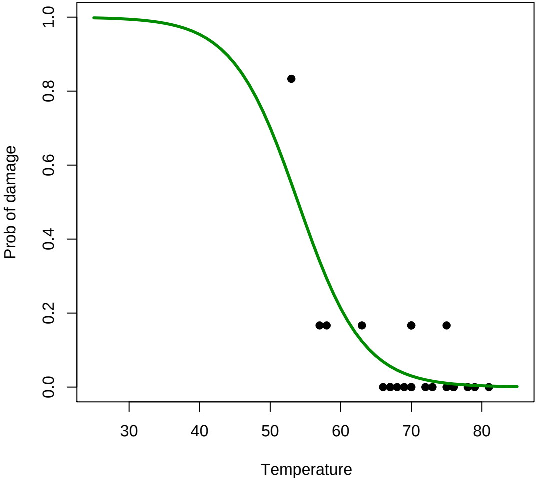
Also, does this look like it meets the Gauss-Markov?

Heteroskedasticity



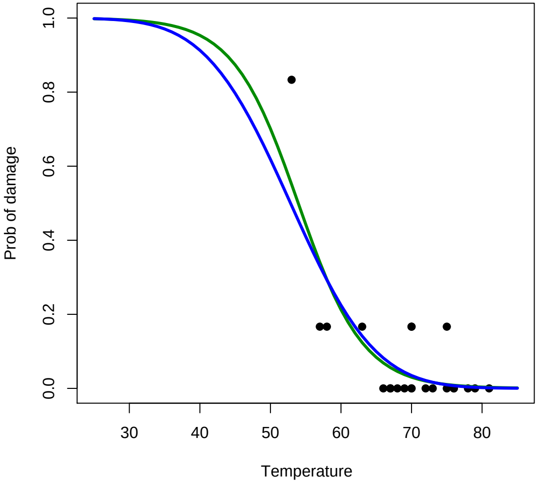
Also, does this look like it meets the Gauss-Markov? NO! (non-constant regression error w.r.t X)

A Better Model: Logistic Regression



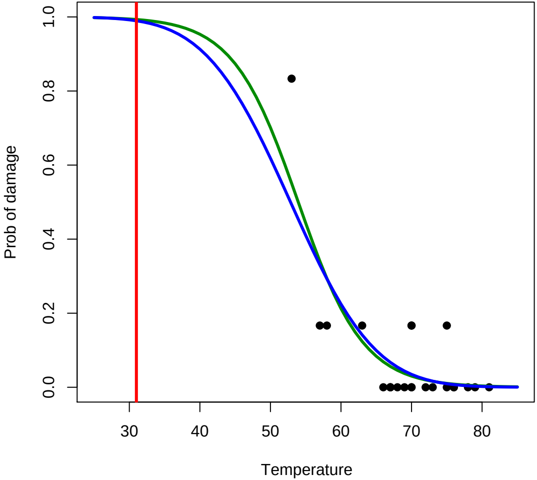
The asymptotic trend towards 0 and 1 (prediction never reaches 0 or 1)

Logit and Probit



Green for logit and blue for probit

Should They Have Launched at $31^\circ F$?



$p(failure) = 0.9930414$ in logit
$p(failure) = 0.9896029$ in probit

Building Binary Outcome Models

- The binary outcome Y_i for $i = 1, \ldots, n$ from the Bernoulli distribution:

$$Pr(Y_i) = p^{y_i}(1 - p)^{1-y_i}$$

- Assume that all Y_i are *independently* and *identically* distributed

The Systematic Component

- A linear systematic component can be constructed as

$$\boldsymbol{X\beta} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

- Advantages of the linear form:
    - accommodates quantitative and qualitative (categorical/factor) predictors
    - allows for transformations (log, centering, standardization)
    - allows combinations of the original predictors (interactions)
    - highly flexible but retains interpretability

The Link Function

- However, $\boldsymbol{X}\beta \neq p \quad (p_i \in [0,1]$, whereas $\boldsymbol{X}\beta \in (-\infty, +\infty))$
- We need a **link function** $g(\cdot)$ such that $\boldsymbol{X_i}\beta = g(p_i)$
- $g(\cdot)$ needs to be monotonic such that $g(p)$ and $\boldsymbol{X}\beta$ have one-to-one correspondence.

Different Link Functions

- Common:
  1. Logit: $g(Y) = \log\left(\frac{p}{1-p}\right)$
  2. Probit: $g(Y) = \mathbf{\Phi}^{-1}(p)$, where $\mathbf{\Phi}$ is the cumulative density function of the standard normal distribution $[\mathcal{N}(0,1)]$
- Rare in political science:
  - Complementary Log-Log (cloglog): $g(Y) = log(-log(1-p))$

Summary: Logit and Probit

1. **the outcome variable**:

$$Y \sim p^y(1-p)^{1-y}$$

2. **the systematic component**:

$$\sum_{j=1}^{k} x_j\beta_j = \mathbf{X}\beta$$

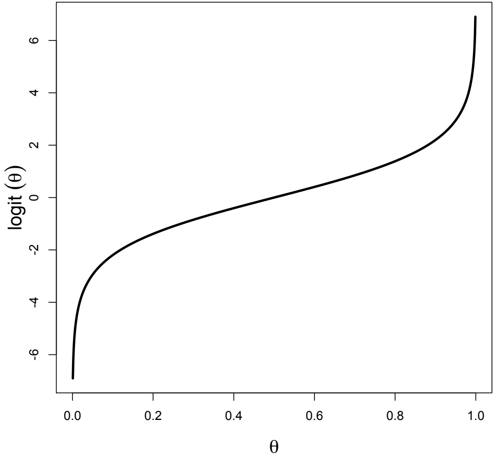
3. **the link function**

$$g(Y) = \ln\left(\frac{p}{1-p}\right) \quad or \quad g(Y) = \mathbf{\Phi}^{-1}(p)$$

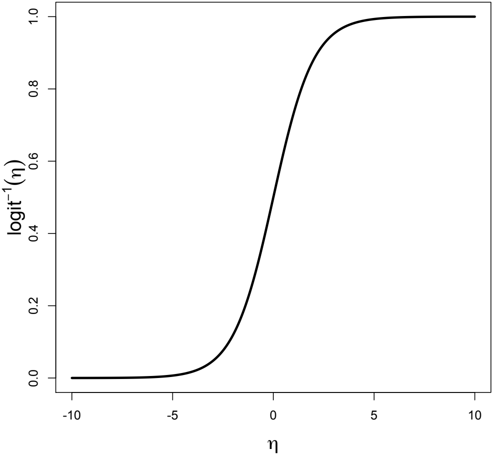
4. **The mean function**

$$p = \frac{\exp(\boldsymbol{X}\boldsymbol{\beta})}{1+\exp(\boldsymbol{X}\boldsymbol{\beta})} \quad or \quad p = \mathbf{\Phi}(\boldsymbol{X}\boldsymbol{\beta})$$

Logit as the Link Function



The Inverse-Logit as the Mean Function



Result: Logit vs. Probit

Table: Probability of graduate school admission

	Logit	Probit
(Intercept)	-4.36^{***}	-2.62^{***}
	(1.04)	(0.61)
gpa	1.05^{***}	0.63^{***}
	(0.30)	(0.18)
AIC	490.97	491.06
BIC	498.95	499.04
Log Likelihood	-243.48	-243.53
Deviance	486.97	487.06
Num. obs.	400	400

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

From $\hat{\beta}$ to $\hat{Pr}$ in Logit

Think about the mean function: $\hat{Pr} = \frac{\exp(X\hat{\beta})}{1+\exp(X\hat{\beta})}$

- Predicted probability is computed by plugging in $X\hat{\beta}$ into the mean function
- Given that the slope of our mean function is always changing, the response of Y to "one-unit change" in X is changing as well:
  - $logit^{-1}(0) = 0.5$, and $logit^{-1}(0.4) = 0.6$: When the sum of the systematic component changes from 0 to 0.4, the corresponding probability changes from 50% to 60% (0.4 unit increase leads to a 10% probability increase)
  - $logit^{-1}(2.2) = 0.9$, and $logit^{-1}(2.6) = 0.93$: When the sum of the systematic component changes from 2.2 to 2.6, the corresponding probability changes from 90% to 93% (0.4 unit increase now leads to only a 3% probability increase).

Four Ways to Interpret Logit Coefficients

1. Compute the predicted probability with the inverse logit function by setting explanatory variables at some values (often at their means).
2. Use the first derivative of the inverse logit function to compute the predicted probability.
3. Divide coefficients by 4 to approximate the maximum marginal effect of a variable ("divided-by-4 rule").
4. Exponentiate coefficients and use the term "odds".

Way 1: Using the Mean Function

- Assuming that your logit model is $logit(y) = -1.40 + 0.33X$.
- The predicted probability can be computed by the inverse logit function: $\frac{\exp(-1.40+0.33X)}{1+\exp(-1.40+0.33X)}$
- How a change in X from 2 to 3 is going to affect Y?

$$logit^{-1}(-1.40 + 0.33 \cdot 3) - logit^{-1}(-1.40 + 0.33 \cdot 2) = 0.08$$

Conclusion: the probability of $Y = 1$ is going to increase by 8%.

When we have more control variables on the right side of the model, we often set those variables at their mean values to make comparison.

Way 2: Differentiating the Inverse Logit Function

- Rather than consider a discrete change in x, we can compute the derivative of the inverse logit function with respect to the explanatory variable of interest.
- Differentiating the function $logit^{-1}(\eta)$ with respect to x:

$$\frac{\partial}{\partial x} logit^{-1}(\eta) = \frac{\beta e^{\eta}}{(1 + e^{\eta})^2}$$

- If $\eta = -1.40 + 0.33X$ and we are interested in the instantaneous effect of X around the value 3.1:
  - plug 3.1 into $\eta = -1.40 + 0.33X$, so we $\eta = -0.377$
  - compute this first derivative value

$$0.33e^{-0.377}/(1 + e^{-0.377})^2 = 0.0796$$

- **Conclusion**: a small change in X around the value of 3.1 leads to a change in the probability of 8.0%

Way 3: Divide-by-4

- The inverse logistic curve is steepest at its center, where $\boldsymbol{X\beta} = 0$.
- Its first derivative at this point is equal to:

$$\frac{\beta e^0}{(1+e^0)^2} = \frac{\beta}{4}$$

- Thus, $\beta/4$ is the maximum difference in $Pr(y=1)$ corresponding to a unit difference in x.
- In practice, we can take an estimated coefficient and divide it by 4 to get an *upper bound* of the predictive difference corresponding to a unit difference in that variable.
- For example, $P(y=1) = logit^{-1}(-1.40 + 0.33 \cdot x)$, and $0.33/4 = 0.08$: an one-unit difference of in X corresponds to no more than an 8% positive difference in the probability of interest.

Way 4: Using Odds

- Another way to interpret logistic regression coefficients is using the term *odds ratios*
- If two outcomes have the probabilities p and $1 - p$, $\frac{p}{1-p}$ is called the odds
- Odds of 1 is equivalent to a probability of 0.5

Interpreting of Coefficients by Odds

- The original link function:

$$ln(\frac{p}{1-p}) = \boldsymbol{X\beta}$$

- Exponentiating both sides:

$$\frac{p}{1-p} = \exp(\boldsymbol{X\beta}) = \exp(\beta_0) \times \exp(\beta_1 x_1) \cdots \times \exp(\beta_k x_k)$$

- The odds are then multiplied by e^{β} if X increases by 1 unit.
- i.e.: if $\beta = 0.2$, then a unit difference in X corresponds to a multiplicative change of $e^{0.2} = 1.22$ in the odds (in other words, 22% increase in the odds).
- But such interpretation is often un-intuitive to many people.

Interpreting Probit Coefficients

- Probit uses the inverse function of the standard normal distribution.
- β_i represents the resulting change in the z-score for the probability in question from one unit increase in X_i, holding constant all other variables.

Coming Up

- Run logit/probit in `R`
- Walk through 4 ways of interpretation for logit with examples
- We will continue to use exponential functions next week for count (discrete) outcome models