Binary Outcomes & Logistic Regression: Model Specification, Interpretation, and Estimation

Jarvis T. Chen (jarvis@hsph.harvard.edu)

6 October 2025

Harvard T.H. Chan School of Public Health

Learning Objectives

- Define a logistic regression model for describing association between a binary random variable and covariates, and interpret the parameters (and their estimates)
- Recognize the similarities with the linear regression model, as part of a broader class of generalized linear models
- Develop the log likelihood and appreciate that maximizing it involves numerical methods and hence need to review software output to ensure that the algorithm has "worked" (e.g., converged)
- Undertake Wald hypothesis tests based expressed in terms of log odds ratios based on the asymptotic normality of the sampling distribution for the MLE
- Construct confidence intervals for the log odds ratio and hence for the odds ratio, and to obtain predictions for the probability of $Y = 1$ given a value for the covariate vector, $\mathbf{X}$.

Readings

The following short readings complement the material covered in this class (emphasis on "complement" rather than "replicate"):

- Hosmer and Lemeshow: Chapters 1, 2 and 3
- Wooldridge: sections 17.1a, 17.1b, 17.1d

An example

Suppose we have a binary outcome, e.g. Y is presence or absence of heart disease, with

$$Y_i = \begin{cases} 1 \text{ if heart disease} \\ 0 \text{ if no heart disease} \end{cases}$$

We have two predictors:

- X_1: Exercise (0 = No regular exercise, 1 = Regular exercise)
- X_2: Smoking Status (0 = Non-smoker, 1 = Smoker)

We might be interested in how exercise and smoking predict the probability of having heart disease in a sample of older adults.

We already have a regression framework for modeling the expected mean of an outcome Y in relation to covariates in $\mathbf{X}$. Could we use OLS to fit a regression model like

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

What issues might arise when using OLS to fit this model?

$$\mathbb{E}(Y|X_1, X_2) = \pi(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Suppose we fit this model and obtain the following output. How do we interpret the regression estimates for β_0, β_1, and β_2?

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.10850    0.01849   5.867 6.04e-09 ***
X1           0.24369    0.02327  10.474  < 2e-16 ***
X2          -0.13209    0.02513  -5.256 1.80e-07 ***
```

An aside: robust variance estimator in R to deal with heteroscedasticity

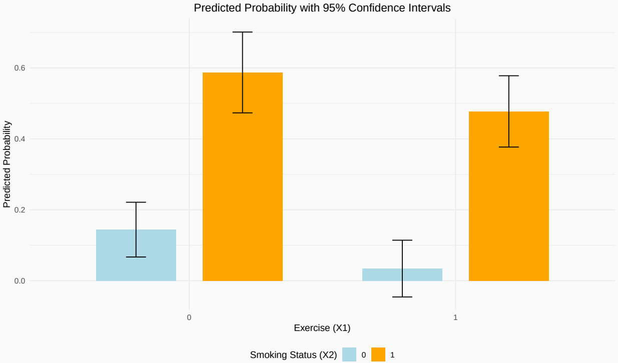
We can use the `sandwich` package in R to get robust standard errors.

```
model <- lm(Y ~ X1 + X2, data = data)

# Use vcovHC to get the robust variance-covariance matrix
robust_vcov <- vcovHC(model, type = "HC1")

# Calculate robust standard errors using HC1
robust_se <- sqrt(diag(robust_vcov))

# Calculate robust confidence intervals
coef_estimates <- coef(lpm_model)  # Extract coefficient estimates
conf_lower <- coef_estimates - 1.96 * robust_se
conf_upper <- coef_estimates + 1.96 * robust_se
```

```
            coef_estimates  conf_lower  conf_upper
(Intercept)      0.1441433  0.06730473  0.22098190
X1              -0.1098287 -0.20885133 -0.01080605
X2               0.4431612  0.30763038  0.57869201
```

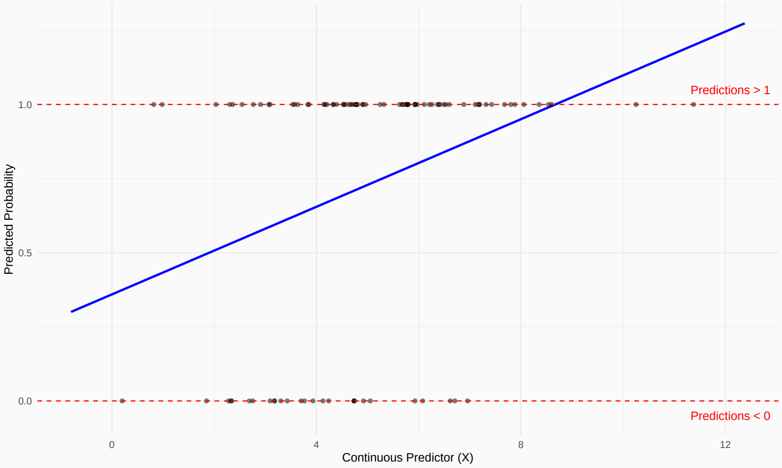
Interpretation of parameter estimates from the Linear Probability Model

- Coefficients represent the additive change in probability (risk) of heart disease for a unit change in the predictors



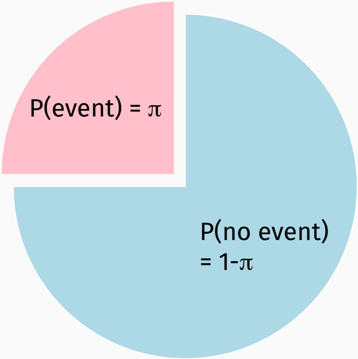
What would happen if we had a continuous X variable?

Linear Probability Model with a continuous covariate



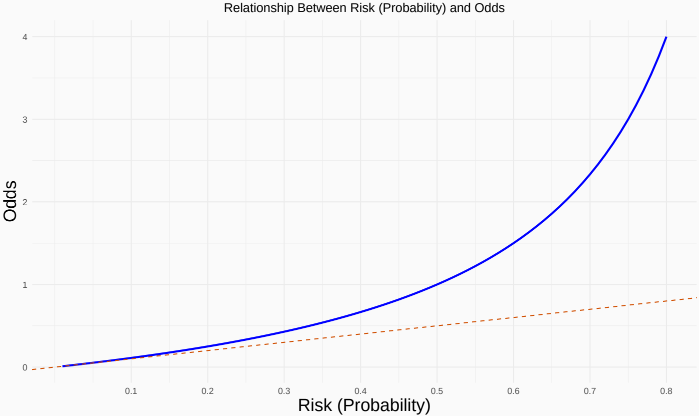
Odds

- $\text{odds} = \frac{\pi}{1-\pi}$, i.e. the ratio of the probability of having an event to the probability of not having an event
- contrast this with the risk, which is the probability of having an event, π
- the odds approximates the risk when π is small (because $1 - \pi \approx 1$)
- Caution: the odds is often misinterpreted as a measure of risk, and the odds ratio incorrectly interpreted as a risk ratio
- only when π is small does the odds ratio approximate the risk ratio!



Probabilities and odds

$$\text{odds} = \frac{\pi}{1-\pi}$$



Notice that while $\pi \in [0, 1]$, $\text{odds} \in [0, \infty)$.

The logit function

What if we take the log of the odds, i.e.

$$\log\left[\frac{\pi}{1-\pi}\right] = \log\left[\frac{Pr(Y=1)}{1-Pr(Y=1)}\right] = \log\left[\frac{\mathbb{E}(Y=1)}{1-\mathbb{E}(Y=1)}\right]$$

Compare

$$\log\left[\frac{\pi}{1-\pi}\right] \in (-\infty, \infty)$$
$$\text{odds} \in [0, \infty)$$
$$\pi \in [0, 1]$$

Note: $\text{logit}(\pi) = \log\left[\frac{\pi}{1-\pi}\right]$

Logistic regression

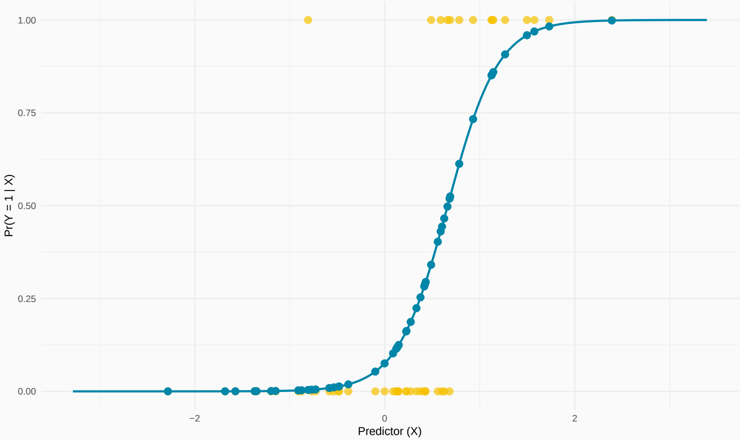
We are interested in $\pi(\mathbf{X}) = \mathbb{E}(\mathbf{X}) = Pr(Y = 1|\mathbf{X})$, i.e. $\pi(\mathbf{X})$ is the probability of the event of interest given $\mathbf{X}$

The **systematic component** of the logistic regression model is

$$\log\left[\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}\right] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Logistic regression

The logit transform "stretches out" the probability scale so that the linear predictor can take on values from $-\infty$ to ∞ while the predicted probability stays in the interval $[0, 1]$



Intuitively, a unit change in X at the extremes of probability (close to 0 or close to 1) results in smaller changes in probability compared with changes closer to 0.5.

Logistic regression as a generalized linear model

In generalized linear model terminology, we are using the logit function as the **link function**; writing $\mu(\mathbf{X}) = \pi(\mathbf{X})$, the full model is:

Systematic component: $\quad \text{logit}\left[\mu(\mathbf{X})\right] = \mathbf{X}^\top \boldsymbol{\beta}$

Random component: $\quad Y|\mathbf{X} \sim \text{Bernoulli}\left[\mu(\mathbf{X})\right]$

Link function: $\quad \text{logit}$

The logit link function facilitates a model that is linear in the parameters.

The expit function

To convert a predicted log odds back into a probability, we use the expit function

$$\pi = \frac{\text{odds}}{1 + \text{odds}}$$

If our logistic regression model is

$$\log\left[\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})}\right] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

then

$$\text{odds}(\mathrm{Y}|\mathbf{X}) = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)$$

so

$$\pi(\mathbf{X}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}$$

Some generalized linear models

Linear regression (normal errors)

- Y is continuous
- Model

$$\mathbb{E}(Y|\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}$$
$$\mu(\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}$$

- Distributional assumption:

$$Y|\mathbf{X} \sim \mathcal{N}\left[\mu(\mathbf{X}), \sigma^2\right]$$

- Link function: identity

Logistic regression

- Y is binary
- Model

$$\text{logit}\left[\mathbb{E}(Y|\mathbf{X})\right] = \mathbf{X}^\top \boldsymbol{\beta}$$
$$\text{logit}\left[\mu(\mathbf{X})\right] = \mathbf{X}^\top \boldsymbol{\beta}$$

- Distributional assumption:

$$Y|\mathbf{X} \sim \text{Bernoulli}\left[\mu(\mathbf{X})\right]$$

- Link function: logit

Note: The **systematic** part of the model involves some function of the mean of Y given X, $\mathbb{E}\left[Y|\mathbf{X}\right]$, being written in terms of a linear combination of the covariates (linear in the parameters $\boldsymbol{\beta}$). The **random** part of the model is the distributional assumption.

Other link functions

- The logit function is the most commonly used link function for binary outcomes
    - Odds and odds ratios are reasonably easy to interpret (though sometimes misinterpreted as rates or risk, and rate ratios or relative risks)
- There are other link functions occasionally used in practice, which transform values of $\pi(\mathbf{X})$ from the 0 to 1 range to the $-\infty$ to ∞ range e.g.:
    - Probit function : $\Phi^{-1}\left[\pi(\mathbf{X})\right]$ where Φ is the standard normal cumulative distribution function
    - Complementary log log function: $\log\left\{-\log\left[\pi(\mathbf{X})\right]\right\}$
    - Arcsin function (inverse of the sine function): $\sin^{-1}\left[\pi(\mathbf{X})\right]$
    - Interpretations are generally more difficult using these functions

Generalized linear models and software

- Generalized linear models (GLMs): a family of models with a common underpinning in statistical theory
    - Therefore can be fitted using a common numerical algorithm
    - Hence often implemented in software as a family – ”GLM” procedures
- Some examples of software
    - Note that each has some form of model specification (about Y and X variables), family/distribution, and link function

      R : `glm(formula y ..., family=..., link=...)`

      STATA : `glm depvar [indepvars], family(...) link(...)`

      SAS : `proc genmod; model y=.../dist=... link=...`

      Note: SAS also has a ”proc glm” but this is for ”general linear models” with a quantitative dependent variable

Interpretation of parameters

Logistic Model (for mean):

$$\text{logit}\left[\pi(\mathbf{X}\right] = \log\left(\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}\right) = \mathbf{X}^\top\boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

- β_0 is the intercept and describes the logit of the probability of the outcome of interest (for which $Y = 1$), or equivalently the log odds of the outcome of interest, when all covariates, X_1 to X_p, take the value zero. (As noted before, this may not have a meaningful interpretation.)
- β_k describes the difference in log odds of the outcome of interest for a unit difference in X_k holding all other covariates constant
  - Noting that $\log(a/b) = \log(a) - \log(b)$
  - Hence, equivalently, β_k describes the log odds ratio for the outcome of interest associated with a unit difference in X_k, holding all other covariates constant.

Interpretation of parameters (2)

Given the model $\log\left\{\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}\right\} = \mathbf{X}^\top\boldsymbol{\beta}$, we can exponentiate both sides of the equation to obtain the odds:

$$\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})} = \exp\{\mathbf{X}^\top\boldsymbol{\beta}\} = \exp\left\{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p\right\},$$

and so e^{β_k} is the odds ratio for the outcome of interest associated with a unit difference in X_k, holding all other covariates constant.

- These are preferred for reporting results as they are more meaningful to most people than log odds ratios.

Note that if the model is additive in the log odds, then this implies that the model is **multiplicative** on the odds scale.

Predicted probabilities

Apply the expit function to obtain the probability $\pi(\mathbf{X})$:

$$\pi(\mathbf{X}) = \frac{e^{\mathbf{X}^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{X}^\top \boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{X}^\top \boldsymbol{\beta}}}$$

This is useful for making predictions for an individual given their specific covariate pattern (i.e., specific values for the vector $\mathbf{X}$).

Interpretation of parameters (3)

Approach to interpretation of parameters (and to estimates of parameters) regarding log odds in logistic regression follows that for standard linear regression

- Same considerations for quantitative, ordinal, binary, and categorical variables
- Same considerations for statistical interactions/effect measure modification

The difference here is that we have used a logit link function, and so generally we exponentiate the parameters (antilog) to facilitate interpretation in terms of the odds of outcome of interest (instead of log odds), and as a step to getting predictions for the probability of the outcome of interest

Example (from Hosmer and Lemeshow)

Estimated logistic regression model for the outcome of a baby being "low birthweight" $(Y = 1)$ or "not low birthweight" $(Y = 0)$, in terms of:

- AGE of the mother (years)
- Indicator variable for whether the mother weighed <110 pounds (LWD=1) or not (LWD=0)

$$\text{logit}(\hat{\pi}) = 0.774 - 1.944 \cdot \text{LWD} - 0.080 \cdot \text{AGE} + 0.132(\text{LWD} * \text{AGE})$$

- How do we interpret the intercept?
- The model includes an interaction, so it might be useful to re-express as:
  - Model for mothers with weight ≥ 110 pounds:

$$\text{logit}(\hat{\pi}) = 0.774 - 0.080 \cdot \text{AGE}$$

  - Model for mothers with weight <110 pounds:

$$\text{logit}(\hat{\pi}) = 0.774 - 1.944 - 0.080 \cdot \text{AGE} + 0.132 \cdot \text{AGE} = -1.170 + 0.052 \cdot \text{AGE}$$

  - What do you note about the estimated age associations?

Example (from Hosmer and Lemeshow)

Predictions can be obtained using the formula:

$$\pi(\mathbf{X}) = \frac{e^{\mathbf{X}^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{X}^\top \boldsymbol{\beta}}}$$

e.g.: AGE = 15, LWD = 0:

$$\text{logit}(\hat{\pi}) = 0.774 - 0.080 \times 15 = -0.426$$

$$\hat{\pi} = \frac{e^{-0.426}}{1 + e^{-0.426}} = 0.40$$

Table of predicted probabilities of a low birthweight baby for selected ages by the mother's pre-pregnancy weight category:

Age	Weight <110 pounds	Weight $\geq$110 pounds
15	0.40	0.40
25	0.53	0.23
35	0.66	0.12

Maximum likelihood estimation for logistic regression

The logistic model has:

$$\log\left(\frac{\pi(\mathbf{X}_i)}{1-\pi(\mathbf{X}_i)}\right) = \mathbf{X}_i^\top \boldsymbol{\beta}$$

and

$$Y|\mathbf{X}_i \sim \text{Bernoulli}[\pi(\mathbf{X}_i)]$$

Assume $Y_i|\mathbf{X}_i$ for $i = 1, \ldots, n$ are independent identically distributed (i.i.d.) with the probability function given by:

$$P(Y_i = 1|\mathbf{X}_i) = \pi(\mathbf{X}_i) \quad \text{and} \quad P(Y_i = 0|\mathbf{X}_i) = 1 - \pi(\mathbf{X}_i)$$

This can be written as:

$$P(Y_i = y_i|\mathbf{X}_i) = [\pi(\mathbf{X}_i)]^{y_i}[1-\pi(\mathbf{X}_i)]^{1-y_i}$$

So the contribution to the likelihood function of a single observation i can be written as:

$$[\pi(\mathbf{X}_i)]^{y_i}[1-\pi(\mathbf{X}_i)]^{1-y_i}$$

Maximum likelihood estimation for logistic regression

Using the iid assumption, the likelihood function for n observations for the vector π, with elements $\pi_i = \pi(x_i)$ for $i = 1, \ldots, n$, is:

$$L(\pi) = \prod_{i=1}^{n} [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

The log-likelihood function ℓ is:

$$\begin{aligned}
\ell(\pi) &= \sum_{i=1}^{n} \log\left([\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i} \right) \\
&= \sum_{i=1}^{n} \{ y_i \log[\pi(x_i)] + (1 - y_i) \log[1 - \pi(x_i)] \}
\end{aligned}$$

The logistic model has $\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = x_i^\top \beta$ and so $\pi(x_i) = \frac{e^{x_i^\top \beta}}{1+e^{x_i^\top \beta}}$. Substituting this into the log likelihood gives us

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i \log\left[\frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}} \right] + (1 - y_i) \log\left[1 - \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}} \right] \right)$$

Approach for developing the log likelihood for GLMs

1. **Using the random component of the model**, i.e., the distributional assumption, write the likelihood function for a single observation
   - i.e., $L(\pi(\mathbf{x}_i))$ for logistic regression
2. **Using the independently identically distributed (iid) assumption**, multiply these together for the n observations, then take the logarithm to obtain the log likelihood
   - i.e., $\ell(\pi(\mathbf{x}_i))$
3. **Rework the systematic component of the model**, i.e., the assumed model for the mean, to express the mean as a function of $\mathbf{x}_i^\top \beta$
   - i.e., $\pi(\mathbf{x}_i) = \frac{e^{\mathbf{x}_i^\top \beta}}{1+e^{\mathbf{x}_i^\top \beta}}$ for logistic regression
4. **Insert this mean into the log likelihood function** to obtain $\ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})$

This approach can be applied to GLMs in general.

Estimation: maximizing the log likelihood

$$\ell(\beta) = \sum_{i=1}^{n} \left(y_i \log \left[\frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}} \right] + (1 - y_i) \log \left[1 - \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}} \right] \right)$$

- Maximize the log likelihood to find the value of $\boldsymbol{\beta}$ which is most likely given the data (i.e., the values of x_i and y_i for $i = 1, \ldots, n$ from a study).
  - Differentiate to obtain $\frac{\partial \ell}{\partial \beta_k}$ for each parameter β_k for $k = 0, 1, \ldots, p$.
  - Set each $\frac{\partial \ell}{\partial \beta_k} = 0$ for $k = 0, 1, \ldots, p$, giving $p + 1$ equations involving $p + 1$ parameters.
  - Solve these equations to find the MLE, $\hat{\beta}$, which maximizes the log likelihood (and hence also the likelihood).

Problem: For logistic regression (and many GLMs and other statistical models), there is **no closed-form analytical solution** to maximizing the likelihood.

Solution: Numerical maximization (or equivalently solving $p + 1$ simultaneous equations) – an iterative process that is implemented in standard software.

Numerical maximization of the likelihood (1)

- For maximum likelihood in general (not just logistic regression), software often uses a numerical maximization algorithm based on the Newton-Raphson Algorithm (aka Newton-Raphson Method).
- For GLMs, an alternative approach used by some software is iteratively reweighted least squares.
    - Recall that MLE and OLS give the same estimator for standard linear regression.
    - For the family of models called GLMs, the distributional assumption can be written in a common format in the **exponential family** of distributions.
    - Iteratively reweighted least squares is essentially an extension of OLS for GLMs that is equivalent to MLE for GLMs.
    - It provides a method for numerically estimating parameters in any GLM (historically very important for software development for statistical models).

Numerical maximization of the likelihood

As a user of software, you should be aware that numerical methods stop when one or more criteria are met:

- e.g., a derivative is “sufficiently close” to zero, or the change in a parameter estimate from one iteration to the next is “sufficiently small.”
- For simple models like logistic regression, good software has default criteria that are generally effective.
- Software often has options for you to control these criteria, which can be important when using more complex models.
- However, software may run into problems and stop without finding the maximum of the likelihood. So, always check for “convergence” and warnings or errors indicating that the numerical algorithm may have failed.
- Common problems include:
    - **Scaling issues**: Very different ranges of measurements for different covariates.
    - **Multicollinearity**: Highly correlated covariates.

Example: logit(blood pressure > 130mmHg) = $\beta_0 + \beta_1 \times BMI$

```
Call:
glm(formula = d.BP$hbp ~ d.BP$bmi, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6368  -0.7917  -0.6881   1.2508   1.9880

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.631071   0.143907. -18.28 <2e-16 ***
d.BP$bmi     0.056643   0.004872   11.63 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4939 on 4297 degrees of freedom
Residual deviance: 4799 on 4295 degrees of freedom
AIC: 4803

Number of Fisher Scoring iterations: 4
```

Example: logit(blood pressure > 130mmHg) = $\beta_0 + \beta_1 \times BMI$

```
Call:
glm(formula = d.BP$hbp ~ d.BP$bmi, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6368  -0.7917  -0.6881   1.2508   1.9880

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.631071   0.143907. -18.28 <2e-16 ***
d.BP$bmi     0.056643   0.004872   11.63 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4939 on 4297 degrees of freedom
Residual deviance: 4799 on 4295 degrees of freedom
AIC: 4803

Number of Fisher Scoring iterations: 4
```

- Generalized linear model
- Mean model specification: outcome variable and covariate
- Distributional assumption and link function (note: we introduced the likelihood through the Bernoulli distribution, which is a special case of the binomial distribution)

Example: logit(blood pressure > 130mmHg) = $\beta_0 + \beta_1 \times BMI$

```
Call:
glm(formula = d.BP$hbp ~ d.BP$bmi, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6368  -0.7917  -0.6881   1.2508   1.9880

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.631071   0.143907. -18.28 <2e-16 ***
d.BP$bmi     0.056643   0.004872   11.63 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4939 on 4297 degrees of freedom
Residual deviance: 4799 on 4295 degrees of freedom
AIC: 4803

Number of Fisher Scoring iterations: 4
```

- The iterations here are about the numerical/computational algorithm used to find the parameter estimates. The number of iterations is usually small. A very large number may be indicative of problems – e.g. if the likelihood function is flattish in the parameter space around the maximum.

Example: logit(blood pressure > 130mmHg) = $\beta_0 + \beta_1 \times BMI$

```
Call:
glm(formula = d.BP$hbp ~ d.BP$bmi, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6368  -0.7917  -0.6881   1.2508   1.9880

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.631071   0.143907. -18.28 <2e-16 ***
d.BP$bmi     0.056643   0.004872   11.63 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4939 on 4297 degrees of freedom
Residual deviance: 4799 on 4295 degrees of freedom
AIC: 4803

Number of Fisher Scoring iterations: 4
```

- Parameter estimates for β_0 and β_1 (maximum likelihood estimates)

Example: logit(blood pressure > 130mmHg) = $\beta_0 + \beta_1 \times BMI$

```
Call:
glm(formula = d.BP$hbp ~ d.BP$bmi, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6368  -0.7917  -0.6881   1.2508   1.9880

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.631071   0.143907. -18.28 <2e-16 ***
d.BP$bmi     0.056643   0.004872   11.63 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4939 on 4297 degrees of freedom
Residual deviance: 4799 on 4295 degrees of freedom
AIC: 4803

Number of Fisher Scoring iterations: 4
```

- Standard errors of the parameter estimates for β_0 and β_1 (maximum likelihood estimates). These are obtained by taking the square roots of the terms on the diagonal of the inverse of the information matrix.

Inferences about β

- Recall: The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ is asymptotically distributed $\mathcal{N}(\boldsymbol{\beta}, V(\hat{\boldsymbol{\beta}}))$, where $V(\hat{\boldsymbol{\beta}})$ is a matrix equal to the inverse of the observed information matrix evaluated at the MLE.
- $V(\hat{\beta}_k)$ is estimated by taking the (k, k)-th diagonal element of $V(\hat{\boldsymbol{\beta}})$, and hence the standard error can be obtained as SE$(\hat{\beta}_k) = \sqrt{V(\hat{\beta}_k)}$
- Note: It can be shown that the observed information matrix can be written as $\mathbf{X}^\top \mathbf{W} \mathbf{X}$, where $\mathbf{X}$ is the $n \times (p+1)$ design matrix containing the covariate data for all subjects (including the intercept), and $\mathbf{W}$ is the diagonal matrix:

$$W = \begin{pmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\pi}_n(1-\hat{\pi}_n) \end{pmatrix}$$

- Hence, $V(\hat{\boldsymbol{\beta}})$ is estimated by $(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}$.
    - Compare this with $\sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$ for standard linear regression.
    - Note that the variance of x's influences the precision in estimating $\boldsymbol{\beta}$, but is also affected by the varying $\hat{\pi}_i(1-\hat{\pi}_i)$ among the individuals.

Hypothesis Testing and Confidence Intervals in Logistic Regression

- The model for the mean is:

$$\text{logit}[\pi(\mathbf{X})] = \log\left(\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}\right) = X^\top \beta$$

- Testing the hypothesis $H_0 : \beta_k = \beta_k^*$ is equivalent to testing the hypothesis:

$$H_0 : e^{\beta_k} = e^{\beta_k^*}$$

  i.e., that the (adjusted) odds ratio associated with a unit difference in covariate x_k equals $e^{\beta_k^*}$.
- **Special case**: $H_0 : \beta_k = 0$ is used for testing the hypothesis that the (adjusted) log odds ratio is 0, which is equivalent to testing the hypothesis that the odds ratio is 1.
- If we calculate a confidence interval for β_k, i.e., for the log odds ratio associated with a one-unit difference in x_k, given by (β_k^L, β_k^U), then $(e^{\beta_k^L}, e^{\beta_k^U})$ is a confidence interval for the odds ratio associated with a one-unit difference in x_k.
- A similar concept applies for other GLMs.

Example: logit(blood pressure > 130mmHg) = $\beta_0 + \beta_1 \times BMI$

```
Call:
glm(formula = d.BP$hbp ~ d.BP$bmi, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6368  -0.7917  -0.6881   1.2508   1.9880

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.631071   0.143907. -18.28 <2e-16 ***
d.BP$bmi     0.056643   0.004872   11.63 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4939 on 4297 degrees of freedom
Residual deviance: 4799 on 4295 degrees of freedom
AIC: 4803

Number of Fisher Scoring iterations: 4
```

- Z-value (and associated p-value) are for test that true parameter is zero, and are obtained based on asymptotic normality of sampling distribution (sample size in this dataset was very large).

Odds ratios and confidence intervals

Parameter estimate for BMI: 0.056643 (standard error = 0.004872).

- 95% CI (two-sided) based on asymptotic normality of the sampling distribution is:

  $$0.056643 \pm z_{1-0.5/2} \times 0.004872 = 0.056643 \pm 1.96 \times 0.004872 = 0.047094 \text{ to } 0.066192$$

- **Interpretation**: Comparing people whose BMI is higher by 1 kg/m^2 than other people, the log odds of high blood pressure is increased by 0.057 (95

Easier interpretation if we transform (take exponentials) back to odds ratios:

- **Odds ratio**: $\exp(0.056643) = 1.058278$
- 95% CI: $\exp(0.047094) = 1.048221$ and $\exp(0.066192) = 1.068432$
- **Interpretation**: Comparing people whose BMI is higher by 1 kg/m^2 than other people, the odds of high blood pressure is increased by a factor of 1.058 (95% CI: 1.048 to 1.068).
- Or: Comparing people whose BMI is higher by 1 kg/m^2, the odds of high blood pressure is increased by 5.8% (95% CI: 4.8
- Or: **Odds ratio** for high blood pressure comparing people whose BMI is higher by 1 kg/m^2 is 1.058 (95% CI: 1.048 to 1.068).

Odds ratios and confidence intervals

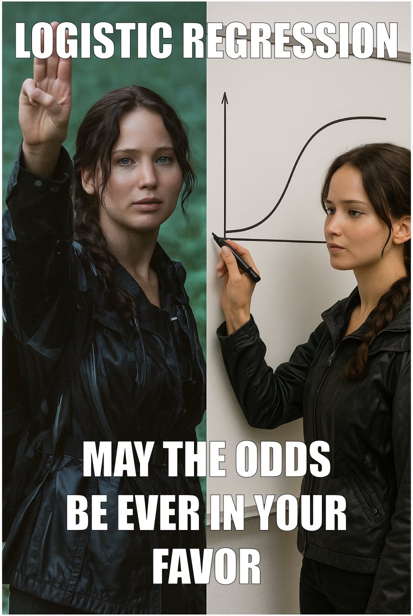
Perhaps better interpretation if we consider BMI differences of 5 kg/m²:

- **log(OR)** = $5 \times 0.056643 = 0.283215$ (95% CI: 5×0.047094 to $5 \times 0.066192 = 0.23547$ to 0.33096).
- **Odds ratio**: 1.327390 (95% CI: 1.265503 to 1.392304).
- Note: OR is $(1.058278)^5$.
- **Interpretation**:
  - Comparing people whose BMI is higher by 5 kg/m² than other people, the odds of high blood pressure is increased by 33% (95% CI: 27% to 39%).
  - Or: **Odds ratio** for high blood pressure comparing people whose BMI is higher by 5 kg/m² than other people is 1.33 (95% CI: 1.27 to 1.39).

Odds ratios and confidence intervals

What if we wanted to compare people whose BMI is lower by 5 kg/m2 than other people?

- **Log OR** = $-5 \times 0.056643 = -0.283215$ (95% CI: -5×0.047094 to $-5 \times 0.066192 = -0.23547$ to -0.33096).
- **Odds ratio**: 0.753358 (95% CI: 0.718234 to 0.790199).
- **Interpretation**:
    - Comparing people whose BMI is lower by 5 kg/m² than other people, the odds of high blood pressure is decreased by 25% (95% CI: 21% to 28%).
    - Or: **Odds ratio** for high blood pressure comparing people whose BMI is lower by 5 kg/m² than other people is 0.75 (95% CI: 0.72 to 0.79).
- **Note**: There is an asymmetry in odds ratios depending on whether we think about BMI being 5 kg/m² higher or lower.
    - Comparing the odds of high blood pressure for those whose BMI is **higher** by 5 kg/m²: odds *increases* by 33%.
    - Comparing the odds of high blood pressure for those whose BMI is **lower** by 5 kg/m²: odds *decreases* by 25%.


LOGISTIC REGRESSION
MAY THE ODDS
BE EVER IN YOUR
FAVOR