# Chapter 3: Logistic Regression Model

**Noryanti Muhammad**
**Centre for Mathematical Sciences**
**College of Computing and Applied Sciences**
**Universiti Malaysia Pahang**

**Centre of Excellence (CoE) for Data Science & Artificial Intelligence**
**Research & Innovation Department**
**Universiti Malaysia Pahang**

# Expected Outcomes:

## By the end of this chapter, students should be able:

- ✓To understand the basic ideas behind modeling categorical data with binary logistic regression.
- ✓To understand how to fit the model and interpret the parameter estimates, especially in terms of odds and odd ratios.
- ✓To investigate the model performance based on Bayesian information criteria (BIC) or Schwarz Bayesian Criteria (SBC) and Akaike information criteria (AIC).

# Content:

# 3.1 Regression with Binary Response Variable

- Linear regression is only dealing with continuous variables instead of Bernoulli variables.

- The problem of Linear Regression is that these predictions are not sensible for classification since the true probability must fall between 0 and 1 but it can be larger than 1 or smaller than 0.

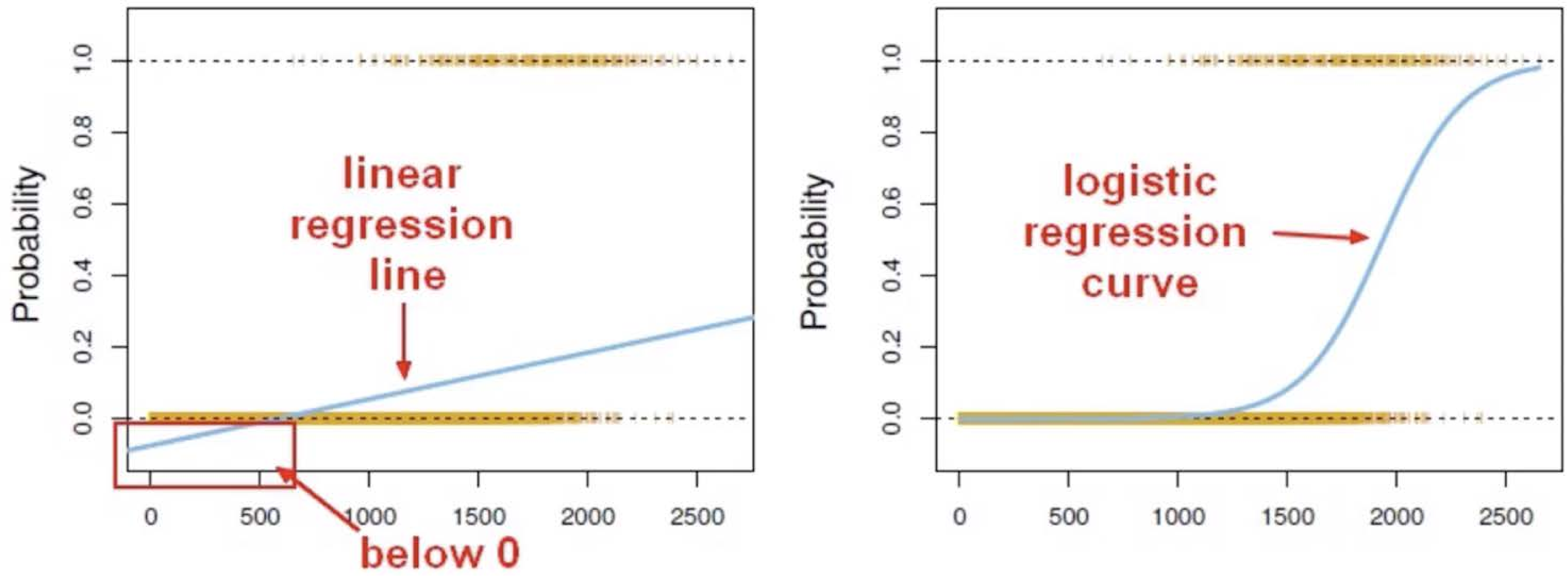
- Noted that classification is not normally distributed which is violated assumption of **Normality.**

- Moreover, both mean and variance depend on the underlying probability.

- Any factor that affects the probability will change not just the mean but also the variance of the observations which means the variance is no longer constantly violating the assumption of **Homoscedasticity**.

- Therefore, we cannot directly apply linear regression because it won't be a good fit.

# 3.1 Regression with Binary Response Variable



# 3.1 Regression with Binary Response Variable

- Generalised linear model (GLM) caters to these situations by allowing for response variables that have arbitrary distributions (other than only normal distributions),
- and by using a **link function** to vary linearly with the predicted values rather than assuming that the response itself must vary linearly with the predictor.
- GLM offers **extra flexibility** in modelling.
- We could apply the **logistic function** (also called the ‘**inverse logit**’ or ‘**sigmoid function**’).

# 3.1 Regression with Binary Response Variable

Before we dig deep into logistic regression, we need to clear up some of the fundamentals of statistical terms — **Probablility** and **Odds**.

The **probability** that an event will occur is the fraction of times you expect to see that event in many trials.

If the probability of an event occurring is Y, then the probability of the event not occurring is 1-Y. Probabilities always range between 0 and 1.

The **odds** are defined as the probability that the event will occur divided by the probability that the event will not occur.

Unlike probability, the odds are not constrained to lie between 0 and 1, but can take any value from zero to infinity.

# 3.1 Regression with Binary Response Variable

If the probability of Success is P, then the odds of that event is:

$$odds = \frac{P}{1 - P}$$

| *Y* | *1* | *0* |
|---|---|---|
| *Pr(Y=1)* | *P* | *1- P* |

**P= Success, 1-P = Failure*

Example:
If the probability of success (P) is 0.60 (60%), then the probability of failure(1-P) is 1–0.60 = 0.40(40%). Then the odds are 0.60 / (1–0.60) = 0.60/0.40 = 1.5.

# 3.1 Regression with Binary Response Variable

**To transform the model from linear regression to logistic regression using the logistic function.**

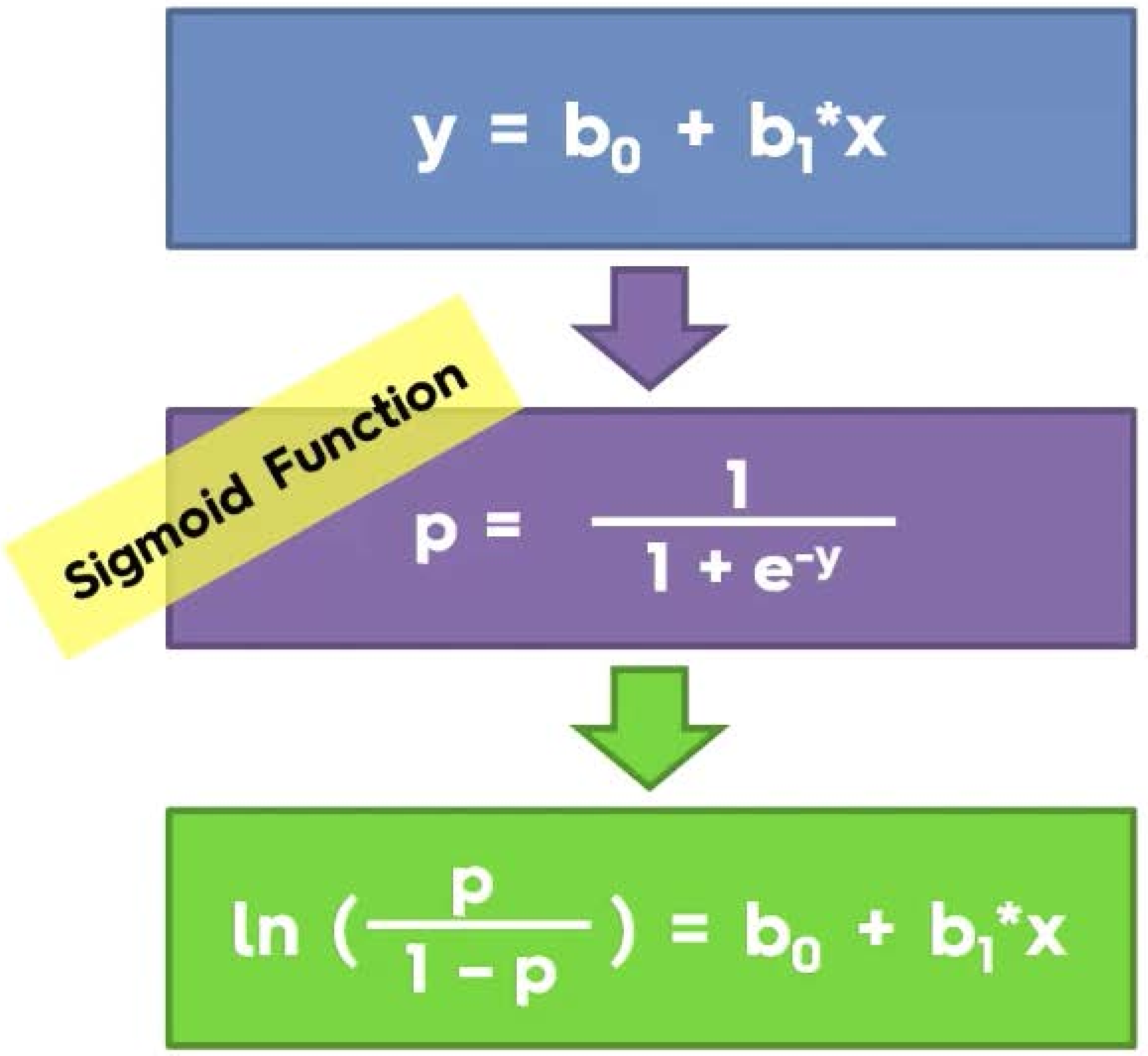
$$\ln\left(\frac{P}{1-P}\right) = b_0 + b_1 x$$

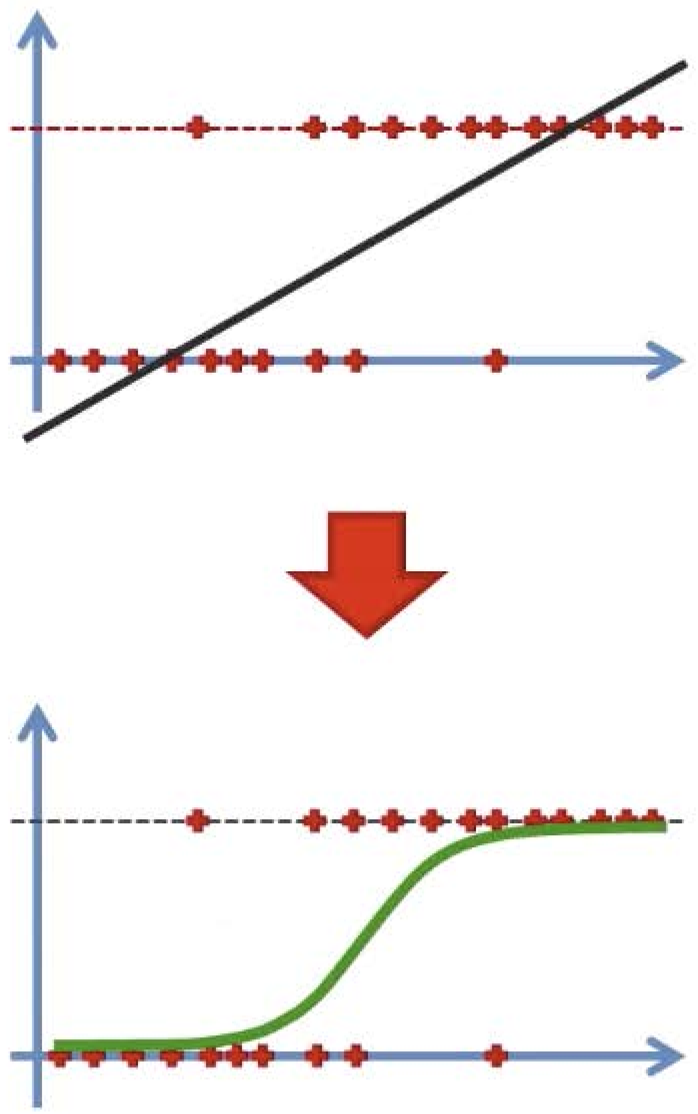
$$\frac{P}{1-P} = e^{b_0 + b_1 x}$$

$$P = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

In (odd)=$b_0$+$b_1$x

logistic function (also called the 'inverse logit').



Logistic Regression is all about predicting binary variables, not predicting continuous variables.

The linear regression is passed through a sigmoid function (logit function) that can map any real value between 0 and 1

# Example

Example 1

Suppose that we are interested in the factors that influence whether a political candidate wins an election. The outcome (response) variable is binary (0/1); win or lose. The predictor variables of interest are the amount of money spent on the campaign, the amount of time spent campaigning negatively and whether or not the candidate is an incumbent.

Example 2.

A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/don't admit, is a binary variable.

# Example

This dataset has a **binary response** (outcome, dependent) variable called admit. There are three predictor variables: gre, gpa and rank.

We will treat the variables gre and gpa as continuous.

The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest.

We can get **basic descriptive** for the entire data set by using **summary**.

To get the **standard deviations**, we use **sapply** to apply the sd function to each variable in the dataset.

# Example

#The code below estimates a logistic regression model using the glm (generalized linear model) function.
#First, we convert rank to a factor to indicate that rank should be treated as a categorical variable.
mydata$rank <- factor(mydata$rank)
mylogit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
summary(mylogit)

#Confidence Interval for the coefficient estimates
## CIs using profiled log-likelihood
confint(mylogit)

## CIs using standard errors
confint.default(mylogit)

**NOTES:**

One useful approach to maximization is to set β1 to a given value and then find the value of β0 that maximizes the log-likelihood g(β) given that value of β1

Profile likelihood is often used when accurate interval estimates are difficult to obtain using standard methods—for example, when the log-likelihood function is highly nonnormal in shape or when there is a large number of nuisance parameters

```
> ## CIs using profiled log-likelihood
> confint(mylogit)
Waiting for profiling to be done...
                    2.5 %       97.5 %
(Intercept) -6.2716202334 -1.792547080
gre          0.0001375921  0.004435874
gpa          0.1602959439  1.464142727
rank2       -1.3008888002 -0.056745722
rank3       -2.0276713127 -0.670372346
rank4       -2.4000265384 -0.753542605
> ## CIs using standard errors
> confint.default(mylogit)
                    2.5 %       97.5 %
(Intercept) -6.2242418514 -1.755716295
gre          0.0001202298  0.004408622
gpa          0.1536836760  1.454391423
rank2       -1.2957512650 -0.055134591
rank3       -2.0169920597 -0.663415773
rank4       -2.3703986294 -0.732528724
```

```
Call:
glm(formula = admit ~ gre + gpa + rank, family = "binomial",
    data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6268  -0.8662  -0.6388   1.1490   2.0790

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500 0.000465 ***
gre          0.002264   0.001094   2.070 0.038465 *
gpa          0.804038   0.331819   2.423 0.015388 *
rank2       -0.675443   0.316490  -2.134 0.032829 *
rank3       -1.340204   0.345306  -3.881 0.000104 ***
rank4       -1.551464   0.417832  -3.713 0.000205 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 499.98  on 399  degrees of freedom
Residual deviance: 458.52  on 394  degrees of freedom
AIC: 470.52

Number of Fisher Scoring iterations: 4
```

# Example

- The **call**, this is R reminding us what the model we ran was, what options we specified, etc.

- Next we see the **deviance residuals**, which are a **measure of model fit**. This part of output shows **the distribution of the deviance residuals for individual cases** used in the model.

How to use summaries of the deviance statistic to assess model fit.

- The next part of the output shows the **coefficients**, their **standard errors**, the **z-statistic** (sometimes called a **Wald z-statistic**), and the associated ***p*-values**. Both gre and gpa are statistically significant, as are the three terms for rank.

- The logistic regression coefficients give **the change in the log odds of the outcome for a one unit increase** in the predictor variable.

  - For every one unit change in gre, the log odds of admission (versus non-admission) increases by 0.002.

  - For a one unit increase in gpa, the log odds of being admitted to graduate school increases by 0.804.

  - The indicator variables for rank have a slightly different interpretation. For example, having attended an undergraduate institution with rank of 2, versus an institution with a rank of 1, changes the log odds of admission by -0.675.

- Below the table of coefficients are **fit indices**, including the null and deviance residuals and the AIC, which use to help assess model fit.

# Example

## Test for an overall effect of rank using the wald.test function

```
#b- coefficient, Sigma -  the variance covariance matrix of the error terms and
#Terms – which terms in R to be tested
wald.test(b = coef(mylogit), Sigma = vcov(mylogit), Terms = 4:6)
```

```
Wald test:
----------

Chi-squared test:
X2 = 20.9, df = 3, P(> X2) = 0.00011
```

The chi-squared test statistic of 20.9, with three degrees of freedom is associated with a p-value of 0.00011 indicating that the overall effect of rank is statistically significant.

# Example

## To exponentiate the coefficients and interpret them as odds-ratios

```
## odds ratios only
exp(coef(mylogit))
## odds ratios and 95% CI
exp(cbind(OR = coef(mylogit), confint(mylogit)))
```

```
> ## odds ratios only
> exp(coef(mylogit))
(Intercept)         gre         gpa       rank2       rank3       rank4
  0.0185001   1.0022670   2.2345448   0.5089310   0.2617923   0.2119375
> ## odds ratios and 95% CI
> exp(cbind(OR = coef(mylogit), confint(mylogit)))
Waiting for profiling to be done...
                   OR       2.5 %    97.5 %
(Intercept) 0.0185001 0.001889165 0.1665354
gre         1.0022670 1.000137602 1.0044457
gpa         2.2345448 1.173858216 4.3238349
rank2       0.5089310 0.272289674 0.9448343
rank3       0.2617923 0.131641717 0.5115181
rank4       0.2119375 0.090715546 0.4706961
>
```

For a one unit increase in gpa, the odds of being admitted to graduate school (versus not being admitted) increase by a factor of 2.23

# Example

## Use predicted probabilities to help you understand the model

```
## create new data
newdata1 <- with(mydata, data.frame(gre = mean(gre), gpa = mean(gpa), rank = factor(1:4)))

#OR the values of rankP should be predictions made using the predict( ) function.
newdata1$rankP <- predict(mylogit, newdata = newdata1, type = "response")
```

```
>
> ## create new data
> newdata1 <- with(mydata, data.frame(gre = mean(gre), gpa = mean(gpa), rank = factor(1:4)))
> newdata1
    gre    gpa rank
1 587.7 3.3899    1
2 587.7 3.3899    2
3 587.7 3.3899    3
4 587.7 3.3899    4
> #OR the values of rankP should be predictions made using the predict( ) function.
> newdata1$rankP <- predict(mylogit, newdata = newdata1, type = "response")
> #OR the values of rankP should be predictions made using the predict( ) function.
> newdata1$rankP
[1] 0.5166016 0.3522846 0.2186120 0.1846684
> newdata1
    gre    gpa rank     rankP
1 587.7 3.3899    1 0.5166016
2 587.7 3.3899    2 0.3522846
3 587.7 3.3899    3 0.2186120
4 587.7 3.3899    4 0.1846684
>
```

The predicted probability of being accepted into a graduate program is 0.52 for students from the highest prestige undergraduate institutions (rank=1), and 0.18 for students from the lowest ranked institutions (rank=4).

# Example

Simulations

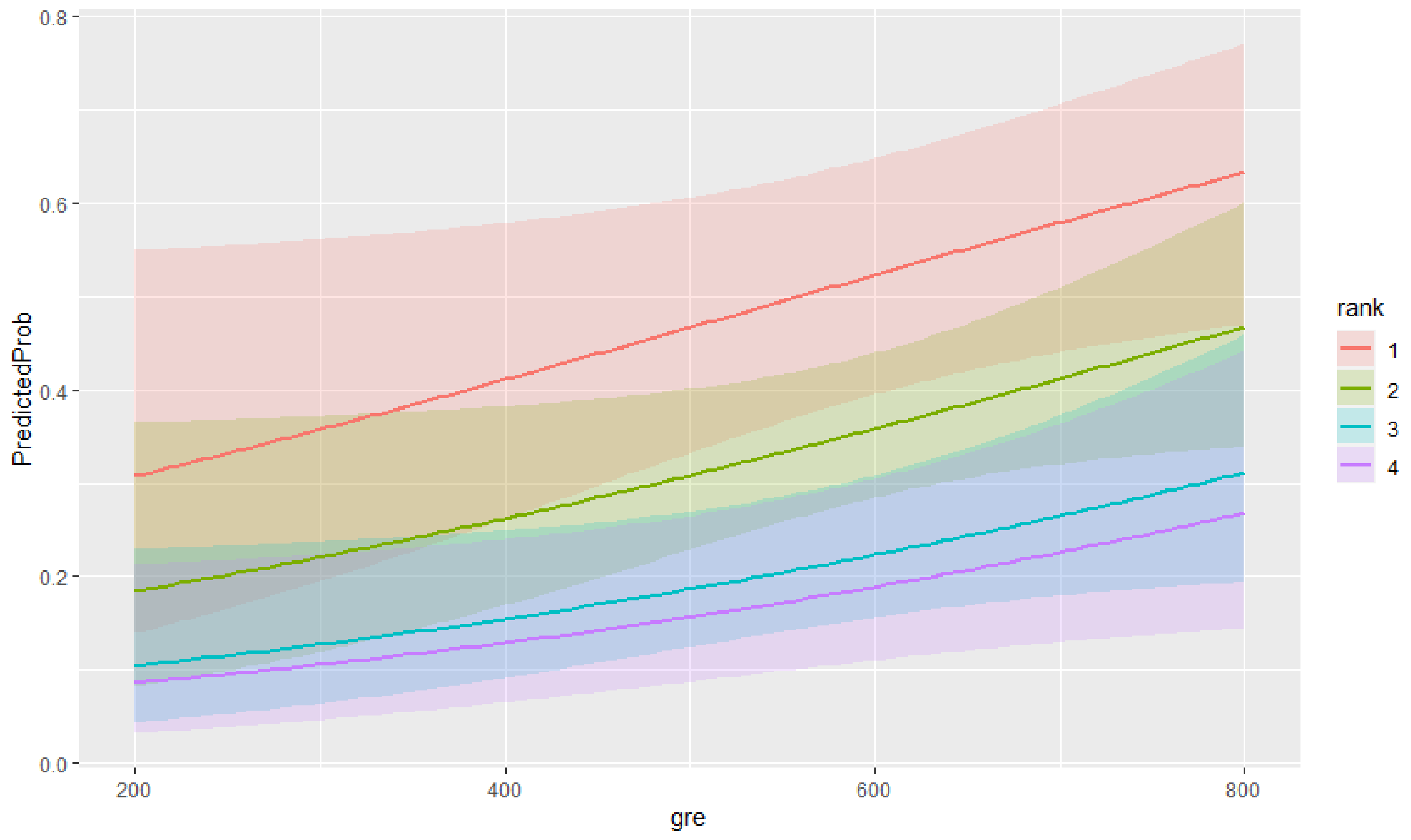
Create 100 values of gre between 200 and 800, at each value of rank (i.e., 1, 2, 3, and 4). Then, add the rank probability as previous slide.

```
#To plot and Create more data
newdata2 <- with(mydata, data.frame(gre = rep(seq(from = 200, to = 800,
length.out = 100), 4), gpa = mean(gpa), rank = factor(rep(1:4, each = 100))))

#The code to generate the predicted probabilities (the first line below) is the
same as before,
#except we are also going to ask for standard errors so we can plot a
confidence interval. We get the
#estimates on the link scale and back transform both the predicted values and
confidence limits into probabilities.
newdata3 <- cbind(newdata2, predict(mylogit, newdata = newdata2, type =
"link", se = TRUE))
newdata3 <- within(newdata3, {
  PredictedProb <- plogis(fit)
  LL <- plogis(fit - (1.96 * se.fit))
  UL <- plogis(fit + (1.96 * se.fit))
})

## view first few rows of final dataset
head(newdata3)

#to use graphs of predicted probabilities to understand and/or present the model
ggplot(newdata3, aes(x = gre, y = PredictedProb)) + geom_ribbon(aes(ymin =
LL, ymax = UL, fill = rank), alpha = 0.2) + geom_line(aes(colour = rank),
```



# Example

## Measures of how well our model fits (Analysis of Deviance)

- The test asks **whether the model with predictors fits significantly better than a model with just an intercept** (i.e., a null model).
- The **test statistic** is the difference between the residual deviance for the model with predictors and the null model.
- The test statistic is distributed chi-squared with degrees of freedom equal to the differences in degrees of freedom between the current and the null model (i.e., the number of predictor variables in the model).
- From Example: *The chi-square of 41.46 with 5 degrees of freedom and an associated p-value of less than 0.001 tells us that our model as a whole fits significantly better than an empty model.*
- This is sometimes called a **likelihood ratio test** (the **deviance residual is -2*log likelihood**)

# General GLM

**Basic idea GLM to develop a linear model for an appropriate function of the expected value of the response variable.**

General Model

$$y_i = \boldsymbol{X'}_{\boldsymbol{i}}\boldsymbol{\beta} + \boldsymbol{\varepsilon_i}$$

where $y_i = 0{,}1$ $\boldsymbol{x'_i} = [1, x_{i1}, x_{i2,\dots,}x_{ik}]$ $\boldsymbol{\beta'} = [\beta_0, \beta_2, \beta_2, \dots, \beta_k]$

| *Y* | *1* | *0* |
|---|---|---|
| *Pr(Y=1)* | *P* | *1- P* |

**P= Success, 1-P = Failure*

$$E(y_i) = 1(P_i) + 0(1 - P) = P \quad \longrightarrow \quad E(y_i) = \boldsymbol{X'}_{\boldsymbol{i}}\boldsymbol{\beta} = P$$

# Link Functions

**Basic idea GLM to develop a linear model for an appropriate function of the expected value of the response variable.**

- Let $\eta_i$ be the linear predictor defined by

$$\eta_i = g[E(y_i)] = g(\mu_j) = x_i'\beta$$

- Note that the expected response is just

$$E(y_i) = g^{-1}(\eta) = g^{-1}(x_i'\beta)$$

- We call g is the link function.
- Canonical link function

One in which transforms the mean, µ = E(yi), to the natural exponential (location) parameter for the exponential family of distributions (e.g., normal, binomial, Poisson, gamma). The canonical link function is the most commonly used link form in generalized linear models

Link function Linear predictor

$$\ln \lambda_i = b_0 + b_1 x_i$$

$$y_i \sim \mathrm{Poisson}(\lambda_i)$$

Probability distribution

# Link Functions

**Basic idea GLM to develop a linear model for an appropriate function of the expected value of the response variable.**

There are many canonical links function, most common choices of distributions employed with the GLM are

1. The probit link

$$\eta_i = \Phi^{-1}[E(y_i)]$$

where $\Phi$ represent the cumulative standard normal distribution function.

**TABLE 13.8 Canonical Links for the Generalized Linear Model**

| Distribution | Canonical Link |
|---|---|
| Normal | $\eta_i = \mu_i$ (identity link) |
| Binomial | $\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ (logistic link) |
| Poisson | $\eta_i = \ln(\lambda)$ (log link) |
| Exponential | $\eta_i = \frac{1}{\lambda_i}$ (reciprocal link) |
| Gamma | $\eta_i = \frac{1}{\lambda_i}$ (reciprocal link) |

# Link Functions

**Basic idea GLM to develop a linear model for an appropriate function of the expected value of the response variable.**

2. The complementary log-log link

$$\eta_i = \ln\{ln[1 - E(y_i)]\}$$

3. The power family link

$$\eta_i = \begin{cases} E(y_i)^\lambda, & \lambda \neq 0 \\ \ln[E(y_i)] \quad , & \lambda = 0 \end{cases}$$

**Notes:** link function take advantages of the natural distribution of the response which different from transformation.

# 3.2 Parameter Estimation

## For Logistic Regression

- Often if the data or the model been used transformation, use Ordinary Least Square (OLS).
- For GLM (variance of the response not constant) use Weighted Least Squares (WLS)
- Numerical search could be used to compute the Maximum Likelihood Estimates (MLEs)
- However, Iteratively reweight least squares (IRLS) to actually find the MLEs.
- Then, finally often called Maximum likelihood score equation.
- The Newtown-Raphson method is actually used to solve the logistic regression model.

# 3.2 Parameter Estimation

## For Logistic Regression

- Often if the data or the model been used transformation, use Ordinary Least Square (OLS).
- For GLM (variance of the response not constant) use Weighted Least Squares (WLS)
- Numerical search could be used to compute the Maximum Likelihood Estimates (MLEs)
- However, Iteratively reweight least squares (IRLS) to actually find the MLEs.
- Then, finally often called Maximum likelihood score equation.
- The Newtown-Raphson method is actually used to solve the logistic regression model.

# 3.2 Parameter Estimation

As explain in lecture.

# 3.3 Diagnostic Checking

- There are three statistical approaches to estimating how well a given model fits a dataset and how complex the model is. And each can be shown to be equivalent or proportional to each other, although each was derived from a different framing or field of study.
  - Akaike Information Criterion (AIC). Derived from frequentist probability. OR Schwarz information criterion (also SIC, SBC, SBIC)
  - Bayesian Information Criterion (BIC). Derived from Bayesian probability.
  - Minimum Description Length (MDL). Derived from information theory.
- Each statistic can be calculated using the log-likelihood for a model and the data. Log-likelihood comes from Maximum Likelihood Estimation, a technique for finding or optimizing the parameters of a model in response to a training dataset.
- Advanced :
  - Consistent Akaike Information Criterion (CAIC)
  - Hannan_Quinn Information Criterion (HQIC)

# 3.3 Diagnostic Checking



## Bayesian information criterion (BIC)

- Bayesian information criterion (BIC) or Schwarz information criterion (also SIC, SBC, SBIC) is a criterion for model selection among a finite set of models.

- Models with lower BIC are generally preferred.

- The BIC was developed by Gideon E. Schwarz and published in a 1978 where a Bayesian argument is been considered.

$$BIC = k \log n - 2 \log L(\hat{\theta})$$

*n* is the sample size; the number of observations or number of data points you are working with.
*k* is the number of parameters which your model estimates, and θ is the set of all parameters.
$L(\hat{\theta})$ represents the likelihood of the model tested, given your data, when evaluated at maximum likelihood values of θ.

# 3.3 Diagnostic Checking

**Properties**

- The BIC generally penalizes free parameters more strongly than the Akaike information criterion, though it depends on the size of n and relative magnitude of n and k.
- It is independent of the prior.
- It can measure the efficiency of the parameterized model in terms of predicting the data.
- It penalizes the complexity of the model where complexity refers to the number of parameters in the model.
- It is approximately equal to the minimum description length criterion but with negative sign.
- It can be used to choose the number of clusters according to the intrinsic complexity present in a particular dataset.
- It is closely related to other penalized likelihood criteria such as Deviance information criterion (DIC) and the Akaike information criterion (AIC).

The BIC suffers from two main limitations.

- the above approximation is only valid for sample size ($n$) much larger than the number $k$ of parameters in the model.
- the BIC cannot handle complex collections of models as in the variable selection (or feature selection) problem in high-dimension.

# 3.3 Diagnostic Checking

## Akaike information criteria (AIC)

- The Akaike information criterion was formulated by the statistician Hirotugu Akaike.
- AIC estimates the relative amount of information lost by a given model: the less information a model loses, the higher the quality of that model.
- In estimating the amount of information lost by a model, AIC deals with the trade-off between the goodness of fit of the model and the simplicity of the model. In other words, AIC deals with both the risk of overfitting and the risk of underfitting.

$$AIC = 2k - 2\ln L(\hat{\theta})$$

*n* is the sample size; the number of observations or number of data points you are working with.
*k* is the number of parameters which your model estimates, and θ is the set of all parameters.
$L(\hat{\theta})$ represents the likelihood of the model tested, given your data, when evaluated at maximum likelihood values of θ.

# 3.3 Diagnostic Checking

- The table will optionally contain test statistics (and P values) comparing the reduction in deviance for the row to the residuals.
- For models with **known dispersion** (e.g., binomial and Poisson fits) the **chi-squared test** is most appropriate,
- Models with **dispersion estimated by moments** (e.g., gaussian, quasibinomial and quasipoisson fits) the **F test** is most appropriate.
- **Mallows' Cp** statistic is the residual deviance plus twice the estimate of σ^2 times the residual degrees of freedom, which is closely related to AIC (and a multiple of it if the dispersion is known).
- Also, can choose "LRT" and "Rao" for **likelihood ratio tests** and **Rao's efficient score test**. The former is synonymous with "Chisq" (although both have an asymptotic chi-square distribution).

# 3.4 Case study

## Prospective study of Coronary Heart Disease (CHD)

Our primary goal is to investigate the relationship between ‘behavior pattern’ and risk of CHD

| | | |
|---|---|---|
| 1 | age | age, years |
| 2 | ht | height, in |
| 3 | wt | weight, lbs |
| 4 | sbp | systolic blood pressure, mmHg |
| 5 | dbp | diastolic blood pressure, mmHg |
| 6 | chol | cholesterol, mg/dL |
| 7 | ncigs | number of cigarettes smoked per day |
| 8 | behave | behavior type 0/1 = B/A |
| 9 | chd | occurrence of a CHD event during follow-up |
| 10 | type | type of CHD event |
| 11 | time | time post-recruitment of the CHD event, days |

# 3.4 Case study

## Diabetes Example

Efron et al. (2004) introduced the diabetes data set with 442 observations and 11 variables. It is often used as an exemplar data set to illustrate new model selection techniques. The following commands will help you get a feel for the data.

| Patient | AGE $x_1$ | SEX $x_2$ | BMI $x_3$ | BP $x_4$ | Serum measurements $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | Response $y$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 59 | 2 | 32.1 | 101 | 157 | 93.2 | 38 | 4 | 4.9 | 87 | 151 |
| 2 | 48 | 1 | 21.6 | 87 | 183 | 103.2 | 70 | 3 | 3.9 | 69 | 75 |
| 3 | 72 | 2 | 30.5 | 93 | 156 | 93.6 | 41 | 4 | 4.7 | 85 | 141 |
| 4 | 24 | 1 | 25.3 | 84 | 198 | 131.4 | 40 | 5 | 4.9 | 89 | 206 |
| 5 | 50 | 1 | 23.0 | 101 | 192 | 125.4 | 52 | 4 | 4.3 | 80 | 135 |
| 6 | 23 | 1 | 22.6 | 89 | 139 | 64.8 | 61 | 2 | 4.2 | 68 | 97 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 441 | 36 | 1 | 30.0 | 95 | 201 | 125.2 | 42 | 5 | 5.1 | 85 | 220 |
| 442 | 36 | 1 | 19.6 | 71 | 250 | 133.2 | 97 | 3 | 4.6 | 92 | 57 |

# 3.4 Case study

## Fertility Example

Predicting fertility score on the basis of socio-economic indicators.

| | Fertility | Agriculture | Examination | Education | Catholic | Infant.Mortality |
|---|---|---|---|---|---|---|
| Courtelary | 80.2 | 17.0 | 15 | 12 | 9.96 | 22.2 |
| Delemont | 83.1 | 45.1 | 6 | 9 | 84.84 | 22.2 |
| Franches-Mnt | 92.5 | 39.7 | 5 | 5 | 93.40 | 20.2 |
| Moutier | 85.8 | 36.5 | 12 | 7 | 33.77 | 20.3 |
| Neuveville | 76.9 | 43.5 | 17 | 15 | 5.16 | 20.6 |
| Porrentruy | 76.1 | 35.3 | 9 | 7 | 90.57 | 26.6 |
| Broye | 83.8 | 70.2 | 16 | 7 | 92.85 | 23.6 |
| Glane | 92.4 | 67.8 | 14 | 8 | 97.16 | 24.9 |
| Gruyere | 82.4 | 53.3 | 12 | 7 | 97.67 | 21.0 |
| Sarine | 82.9 | 45.2 | 16 | 13 | 91.38 | 24.4 |
| Veveyse | 87.1 | 64.5 | 14 | 6 | 98.61 | 24.5 |
| Aigle | 64.1 | 62.0 | 21 | 12 | 8.52 | 16.5 |
| Aubonne | 66.9 | 67.5 | 14 | 7 | 2.27 | 19.1 |
| Avenches | 68.9 | 60.7 | 19 | 12 | 4.43 | 22.7 |

# Summary

- Categorical data model using Logistic Regression Model.
- For GLM parameter estimation using Weighted Least Squares (WLS) i.e Iteratively reweight least squares (IRLS)
- Model selection and diagnostic checking by using Bayesian information criteria (BIC) or Schwarz Bayesian Criteria (SBC) and Akaike information criteria (AIC)


Thank You!