Logistic Regression & Classification
Logistic Regression Model

In 1938, Ronald Fisher and Frank Yates suggested the logit link for regression with a binary response variable.

\[
\ln(\text{Odds of } Y = 1 \mid x) = \ln \left( \frac{P(Y=1 \mid x)}{P(Y=0 \mid x)} \right) = \ln \left( \frac{P(Y=1 \mid x)}{1 - P(Y=1 \mid x)} \right)
\]

\[
\logit[\pi(x)] = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x
\]
Logistic regression model is the most popular model for binary data.

Logistic regression model is generally used to study the relationship between a binary response variable and a group of predictors (can be either continuous or categorical).

\[ Y = 1 \text{ (true, success, YES, etc.) or } Y = 0 \text{ (false, failure, NO, etc.)} \]

Logistic regression model can be extended to model a categorical response variable with more than two categories. The resulting model is sometimes referred to as the multinomial logistic regression model (in contrast to the ‘binomial’ logistic regression for a binary response variable.)
Consider a binary response variable Y=0 or 1 and a single predictor variable x. We want to model \( E(Y|x) = P(Y=1|x) \) as a function of x. The logistic regression model expresses the logistic transform of \( P(Y=1|x) \) as a linear function of the predictor.

\[
\ln\left(\frac{P(Y = 1|x)}{1-P(Y = 1|x)}\right) = \beta_0 + \beta_1 x
\]

This model can be rewritten as

\[
P(Y = 1|x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x)}
\]

\( E(Y|x) = P(Y=1|x) \times 1 + P(Y=0|x) \times 0 = P(Y=1|x) \) is bounded between 0 and 1 for all values of x. The following linear model may violate this condition sometimes:

\[
P(Y = 1|x) = \beta_0 + \beta_1 x
\]
More reasonable modeling

$P_i$  \hspace{1cm} \text{Logit}(P_i)$

**Predictor**  \hspace{1cm} **Predictor**

Logit Transform
In the simple logistic regression, the regression coefficient $\beta_1$ has the interpretation that it is the log of the odds ratio of a success event ($Y=1$) for a unit change in $x$.

$$\ln\left(\frac{P(Y = 1 \mid x + 1)}{P(Y = 0 \mid x + 1)}\right) - \ln\left(\frac{P(Y = 1 \mid x)}{P(Y = 0 \mid x)}\right) = [\beta_0 + \beta_1(x + 1)] - [\beta_0 + \beta_1 x] = \beta_1$$

For multiple predictor variables, the logistic regression model is

$$\ln\left(\frac{P(Y = 1 \mid x_1, x_2, \ldots, x_k)}{P(Y = 0 \mid x_1, x_2, \ldots, x_k)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$
The Logistic Regression Model

\[ \ln \left( \frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \]

where:

- \( \ln \) is the natural logarithm function.
- \( \pi \) is the probability of a “success” \( (P(y = 1)) \).
- \( \beta_0 \) is the \( y \) intercept.
- \( \beta_k \) is the first order logistic regression coefficient for the \( k \)th predictor.
- \( x_k \) is the value of the \( k \)th predictor.

These can be either quantitative or categorical.
\[ \pi = \mu = E(y) = P(y = 1) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)} \]

where:

\[ \pi = \pi(x) = \mu = \mu(x) \]

- \( E(Y) \) is the expected value of the logit model.
- \( P(y = 1) \) is the probability of a “success”.
- \( \beta_0 \) is the y intercept.
- \( \beta_k \) is the first order logistic regression coefficient for the \( k \)th predictor.
- \( x_k \) is the value of the \( k \)th predictor.

These can be either quantitative or categorical.
Logistic Regression, SAS Procedure

- [http://www.ats.ucla.edu/stat/sas/output/SAS_logit_output.htm](http://www.ats.ucla.edu/stat/sas/output/SAS_logit_output.htm)
- **Proc Logistic**

This page shows an example of logistic regression with footnotes explaining the output. The data were collected on 200 high school students, with measurements on various tests, including science, math, reading and social studies. The response variable is high writing test score (**honcomp**), where a writing score greater than or equal to 60 is considered high, and less than 60 considered low; from which we explore its relationship with gender (**female**), reading test score (**read**), and science test score (**science**). The dataset used in this page can be downloaded from

- [http://www.ats.ucla.edu/stat/sas/webbooks/reg/default.htm](http://www.ats.ucla.edu/stat/sas/webbooks/reg/default.htm)
Logistic Regression, SAS Procedure

```sas
data logit;
set "c:\Temp\hsb2";
honcomp = (write >= 60);
run;

proc logistic data= logit descending;
model honcomp = female read science;
run;

proc export data=logit
outfile = "c:\Temp\hsb2.xls"
dbms = xls replace;
run;
```
Logistic Regression, SAS Procedure

data logit;
set "c:\Temp\hsb2";
honcomp = (write >= 60);
run;
proc logistic data= logit descending;
model honcomp = female read science;
output out=pred p=phat lower = lcl upper = ucl;
run;
proc export data=pred
outfile = "c:\Temp\pred.xls"
dbms = xls replace;
run;
Logistic Regression, SAS Procedure

```sas
data logit;
set "c:\Temp\hsb2";
honcomp = (write >= 60);
run;
proc logistic data= logit descending;
model honcomp = female read science /selection=stepwise;
run;
```
Descending option in proc logistic and proc genmod

- The *descending* option in SAS causes the levels of your response variable to be sorted from highest to lowest (by default, SAS models the probability of the lower category).

- In the binary response setting, we code the event of interest as a ‘1’ and use the *descending* option to model that probability $P(Y = 1 \mid X = x)$.

- In our SAS example, we’ll see what happens when this option is not used.
### Logistic Regression, SAS Output

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Only</th>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td>233.289</td>
<td>168.236</td>
</tr>
<tr>
<td>ICC</td>
<td>236.587</td>
<td>181.430</td>
</tr>
<tr>
<td>Log L</td>
<td>231.289</td>
<td>160.236</td>
</tr>
</tbody>
</table>

#### Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>71.0525</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>58.6092</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>39.8751</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

#### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-12.7772</td>
<td>1.9759</td>
<td>41.8176</td>
<td>&lt;.00</td>
</tr>
<tr>
<td>female</td>
<td>1</td>
<td>1.4825</td>
<td>0.4474</td>
<td>10.9799</td>
<td>0.00</td>
</tr>
<tr>
<td>read</td>
<td>1</td>
<td>0.1035</td>
<td>0.0258</td>
<td>16.1467</td>
<td>&lt;.00</td>
</tr>
<tr>
<td>science</td>
<td>1</td>
<td>0.0948</td>
<td>0.0305</td>
<td>9.6883</td>
<td>0.00</td>
</tr>
</tbody>
</table>

#### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>4.404</td>
<td>1.832 - 10.584</td>
</tr>
<tr>
<td>read</td>
<td>1.109</td>
<td>1.054 - 1.167</td>
</tr>
<tr>
<td>science</td>
<td>1.099</td>
<td>1.036 - 1.167</td>
</tr>
</tbody>
</table>
LOGISTIC and GENMOD procedure for a single continuous predictor

```sas
PROC LOGISTIC DATA= dataset <options>;
   MODEL response=predictor /<options>;
   OUTPUT OUT=SAS-dataset keyword=name </option>;
RUN;
```

```sas
PROC GENMOD DATA=dataset <options>;
   MAKE 'OBSTATS' OUT=SAS-data-set;
   MODEL response=predictors </options>;
RUN;
```

😊 The GENMOD procedure is much more general than the logistic procedure as it can accommodate other generalized linear model, and also cope with repeated measures data. It is not required for exam.

Logistic Regression

- Logistic Regression - Dichotomous Response variable and numeric and/or categorical explanatory variable(s)
  - Goal: Model the probability of a particular outcome as a function of the predictor variable(s)
  - Problem: Probabilities are bounded between 0 and 1
- Distribution of Responses: Binomial
- Link Function:
  \[ g(\mu) = \ln \left( \frac{\mu}{1 - \mu} \right) \]
Logistic Regression with 1 Predictor

- Response - Presence/Absence of characteristic
- Predictor - Numeric variable observed for each case
- Model - $\pi(x) = \text{Probability of presence at predictor level } x$

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

- $\beta_1 = 0 \Rightarrow \text{P(Presence) is the same at each level of } x$
- $\beta_1 > 0 \Rightarrow \text{P(Presence) increases as } x \text{ increases}$
- $\beta_1 < 0 \Rightarrow \text{P(Presence) decreases as } x \text{ increases}$
Hypothesis testing

- Significance tests focuses on a test of $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$.
- The Wald, Likelihood Ratio, and Score test are used (we’ll focus on Wald method)
- Wald CI easily obtained, score and LR CI numerically obtained.
- For Wald, the 95% CI (on the log odds scale) is $\hat{\beta}_1 \pm 1.96(SE(\hat{\beta}_1))$
Logistic Regression with 1 Predictor

- \( \beta_0, \beta_1 \) are unknown parameters and must be estimated using statistical software such as SAS, or R
- Primary interest in estimating and testing hypotheses regarding \( \beta_1 \), where S.E. stands for standard error.
  - Large-Sample test (Wald Test):
    - \( H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0 \)
    - \( T.S.: X_{\text{obs}}^2 = \left( \frac{\hat{\beta}_1}{\text{S.E. } \hat{\beta}_1} \right)^2 \)
    - \( R.R.: X_{\text{obs}}^2 \geq \chi^2_{\alpha,1} \)
    - \( P - \text{val}: P(\chi^2 \geq X_{\text{obs}}^2) \)
Example - Rizatriptan for Migraine

- Response - Complete Pain Relief at 2 hours (Yes/No)
- Predictor - Dose (mg): Placebo (0), 2.5, 5, 10

<table>
<thead>
<tr>
<th>Dose</th>
<th># Patients</th>
<th># Relieved</th>
<th>% Relieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>67</td>
<td>2</td>
<td>3.0</td>
</tr>
<tr>
<td>2.5</td>
<td>75</td>
<td>7</td>
<td>9.3</td>
</tr>
<tr>
<td>5</td>
<td>130</td>
<td>29</td>
<td>22.3</td>
</tr>
<tr>
<td>10</td>
<td>145</td>
<td>40</td>
<td>27.6</td>
</tr>
</tbody>
</table>
## SAS Data

<table>
<thead>
<tr>
<th>Dose</th>
<th># Patients</th>
<th># Relieved</th>
<th>% Relieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>67</td>
<td>2</td>
<td>3.0</td>
</tr>
<tr>
<td>2.5</td>
<td>75</td>
<td>7</td>
<td>9.3</td>
</tr>
<tr>
<td>5</td>
<td>130</td>
<td>29</td>
<td>22.3</td>
</tr>
<tr>
<td>10</td>
<td>145</td>
<td>40</td>
<td>27.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>Dose</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2.5</td>
<td>68</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>7</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>101</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>105</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>40</td>
</tr>
</tbody>
</table>
SAS Procedure

Data drug;
Input Y Dose Number;
Datalines;
0 0 65
1 0 2
0 2.5 68
1 2.5 7
0 5 101
1 5 29
0 10 105
1 10 40
;
Run;

proc logistic data= drug descending;
model Y = Dose;
freq Number;
run;
Example - Rizatriptan for Migraine

<table>
<thead>
<tr>
<th>Variables in the Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

^\pi(x) = \frac{e^{-2.490 + 0.165x}}{1 + e^{-2.490 + 0.165x}}

H_0 : \beta_1 = 0 \quad H_A : \beta_1 \neq 0

T.S. : \chi^2_{obs} = \left(\frac{0.165}{0.037}\right)^2 = 19.819

RR : \chi^2_{obs} \geq \chi^2_{0.05,1} = 3.84

P-val : .000
Interpretation of a single continuous parameter

- The sign ($\pm$) of $\beta_1$ determines whether the log odds of $y$ is increasing or decreasing for every 1-unit increase in $x$.
- If $\beta_1 > 0$, there is an increase in the log odds of $y$ for every 1-unit increase in $x$.
- If $\beta_1 < 0$, there is a decrease in the log odds of $y$ for every 1-unit increase in $x$.
- If $\beta_1 = 0$ there is no linear relationship between the log odds and $x$. 
Odds Ratio

- Interpretation of Regression Coefficient ($\beta_1$):
  - In linear regression, the slope coefficient is the change in the mean response as $x$ increases by 1 unit.
  - In logistic regression, we can show that:

$$\frac{odds(x + 1)}{odds(x)} = e^{\beta_1} \quad \left( odds(x) = \frac{\pi(x)}{1 - \pi(x)} \right)$$

- Thus $e^\beta$ represents the change in the odds of the outcome (multiplicatively) by increasing $x$ by 1 unit.
  - If $\beta_1 = 0$, the odds and probability are the same at all $x$ levels ($e^{\beta_1} = 1$)
  - If $\beta_1 > 0$, the odds and probability increase as $x$ increases ($e^{\beta_1} > 1$)
  - If $\beta_1 < 0$, the odds and probability decrease as $x$ increases ($e^{\beta_1} < 1$)
95% Confidence Interval for Odds Ratio

- Step 1: Construct a 95% CI for $\beta$:

$$
\hat{\beta}_1 \pm 1.96 \times [S.E.\hat{\beta}_1] \equiv \left( \hat{\beta}_1 - 1.96 \times [S.E.\hat{\beta}_1], \hat{\beta}_1 + 1.96 \times [S.E.\hat{\beta}_1] \right)
$$

- Step 2: Raise $e = 2.718$ to the lower and upper bounds of the CI:

$$
\left( e^{\hat{\beta}_1 - 1.96 \times [S.E.\hat{\beta}_1]}, e^{\hat{\beta}_1 + 1.96 \times [S.E.\hat{\beta}_1]} \right)
$$

- If entire interval is above 1, conclude positive association
- If entire interval is below 1, conclude negative association
- If interval contains 1, cannot conclude there is an association
Example - Rizatriptan for Migraine

- 95% CI for $\beta_1$:

$$\hat{\beta}_1 = 0.165 \hspace{1cm} S.E. \hat{\beta}_1 = 0.037$$

95% CI: $0.165 \pm 1.96(0.037) = (0.0925, 0.2375)$

- 95% CI for population odds ratio:

$$\left( e^{0.0925}, e^{0.2375} \right) \equiv (1.10, 1.27)$$

- Conclude positive association between dose and probability of complete relief
Logistic regression model with a single categorical \( (\geq 2 \text{ levels}) \) predictor

\[
\logit(p_k) = \log \text{ (odds)} = \beta_0 + \beta_k X_k
\]

where

- \( \logit(p_k) \) logit transformation of the probability of the event
- \( \beta_0 \) intercept of the regression line
- \( \beta_k \) difference between the logits for category \( k \) vs. the reference category
LOGISTIC and GENMOD procedures for a single categorical predictor

PROC LOGISTIC DATA=dataset <options>;
   CLASS variables </option>;
   MODEL response=predictors </options>;
   OUTPUT OUT=SAS-data-set keyword=name </option>;
RUN;

PROC GENMOD DATA=dataset <options>;
   CLASS variables </option>;
   MAKE ‘OBSTATS’ OUT=SAS-data-set;
   MODEL response=predictors </options>;
RUN;
**Class statement in proc logistic**

- SAS will create dummy variables for a categorical variable if you tell it to.
- We need to specify dummy coding by using the `param = ref` option in the `class` statement; we can also specify the comparison group by using the `ref =` option after the variable name.
- Using `class` automatically generates a test of significance for all parameters associated with the class variable (table of Type 3 tests); if you use dummy variables instead (more on this soon), you will not automatically get an “overall” test for that variable.
- We will see this more clearly in the SAS examples.
Reference category

- Each factor has as many parameters as categories, but one is redundant, so we need to specify a *reference* category.
- Similar concept to what you have just learned in the linear regression analysis.
Interpretation of a single categorical parameter

- If your reference group is level 0, then the coefficient of $\beta_k$ represents the difference in the log odds between level $k$ of your variable and level 0.

- Therefore, $e^{\beta_k}$ is an odds ratio for category $k$ vs. the reference category of $x$. 


Creating your own dummy variables and not using the `class` statement

- An equivalent model uses dummy variables (that you create), which accounts for redundancy by not including a dummy variable for your reference category.
- The choice of reference category is arbitrary.
- Remember, this method will not produce an “overall” test of significance for that variable.
Multiple Logistic Regression

- Extension to more than one predictor variable (either numeric or dummy variables).
- With $k$ predictors, the model is written:

$$\pi = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

- Adjusted Odds ratio for raising $x_i$ by 1 unit, holding all other predictors constant:

$$OR_i = e^{\beta_i}$$

- Many models have nominal/ordinal predictors, and widely make use of dummy variables.
Testing Regression Coefficients

Testing the overall model:

\[ H_0 : \beta_1 = \cdots = \beta_k = 0 \]
\[ H_A : \text{Not all } \beta_i = 0 \]

\[ T.S. \ X_{obs}^2 = (-2 \log(L_0)) - (-2 \log(L_1)) \]

\[ R.R. \ X_{obs}^2 \geq \chi^2_{\alpha,k} \]

\[ P = P(\chi^2 \geq X_{obs}^2) \]

- \( L_0, L_1 \) are values of the maximized likelihood function, computed by statistical software packages. This logic can also be used to compare full and reduced models based on subsets of predictors. Testing for individual terms is done as in model with a single predictor.
Example - ED in Older Dutch Men

- **Response:** Presence/Absence of ED \((n=1688)\)
- **Predictors:** \((p=12)\)
  - Age stratum (50-54*, 55-59, 60-64, 65-69, 70-78)
  - Smoking status (Nonsmoker*, Smoker)
  - BMI stratum (<25*, 25-30, >30)
  - Lower urinary tract symptoms (None*, Mild, Moderate, Severe)
  - Under treatment for cardiac symptoms (No*, Yes)
  - Under treatment for COPD (No*, Yes)
  
  * Baseline group for dummy variables
## Example - ED in Older Dutch Men

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>s_b</th>
<th>Adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 55-59 (vs 50-54)</td>
<td>0.83</td>
<td>0.42</td>
<td>2.3 (1.0 – 5.2)</td>
</tr>
<tr>
<td>Age 60-64 (vs 50-54)</td>
<td>1.53</td>
<td>0.40</td>
<td>4.6 (2.1 – 10.1)</td>
</tr>
<tr>
<td>Age 65-69 (vs 50-54)</td>
<td>2.19</td>
<td>0.40</td>
<td>8.9 (4.1 – 19.5)</td>
</tr>
<tr>
<td>Age 70-78 (vs 50-54)</td>
<td>2.66</td>
<td>0.41</td>
<td>14.3 (6.4 – 32.1)</td>
</tr>
<tr>
<td>Smoker (vs nonsmoker)</td>
<td>0.47</td>
<td>0.19</td>
<td>1.6 (1.1 – 2.3)</td>
</tr>
<tr>
<td>BMI 25-30 (vs &lt;25)</td>
<td>0.41</td>
<td>0.21</td>
<td>1.5 (1.0 – 2.3)</td>
</tr>
<tr>
<td>BMI &gt;30 (vs &lt;25)</td>
<td>1.10</td>
<td>0.29</td>
<td>3.0 (1.7 – 5.4)</td>
</tr>
<tr>
<td>LUTS Mild (vs None)</td>
<td>0.59</td>
<td>0.41</td>
<td>1.8 (0.8 – 4.3)</td>
</tr>
<tr>
<td>LUTS Moderate (vs None)</td>
<td>1.22</td>
<td>0.45</td>
<td>3.4 (1.4 – 8.4)</td>
</tr>
<tr>
<td>LUTS Severe (vs None)</td>
<td>2.01</td>
<td>0.56</td>
<td>7.5 (2.5 – 22.5)</td>
</tr>
<tr>
<td>Cardiac symptoms (Yes vs No)</td>
<td>0.92</td>
<td>0.26</td>
<td>2.5 (1.5 – 4.3)</td>
</tr>
<tr>
<td>COPD (Yes vs No)</td>
<td>0.64</td>
<td>0.28</td>
<td>1.9 (1.1 – 3.6)</td>
</tr>
</tbody>
</table>

Interpretations: Risk of ED appears to be:

- Increasing with age, BMI, and LUTS strata
- Higher among smokers
- Higher among men being treated for cardiac or COPD
Logistic Regression – Variable Selection

- Feature selection
  - Stepwise variable selection
  - Best subset variable selection -- Find a subset of the variables that are sufficient for explaining their joint effect on the response.

- Regularization
  - Maximum penalized likelihood
  - Shrinking the parameters via an $L_1$ constraint, imposing a margin constraint in the separable case

\[
\sum_{i=1}^{n} \log P_\beta(y_i \mid x_i) - \frac{C}{2} \left\| \beta \right\|^2
\]
Predicted Values

The output of a logit model is the predicted probability of a success for each observation.
Predicted Values

These are obtained and stored in a separate SAS data set using the OUTPUT statement (see the following code).

```
DM 'LOG;CLEAR;OUT;CLEAR;';

PROC LOGISTIC DATA = W1.OUTBREAK DESCENDING;
  MODEL DISEASE = AGE / CL;
  OUTPUT OUT = LOUT P=P L=L U=U;
RUN;
```
PROC LOGISTIC outputs the predicted values and 95% CI limits to an output data set that also contains the original raw data.

<table>
<thead>
<tr>
<th>Obs</th>
<th>ID</th>
<th>AGE</th>
<th>SES</th>
<th>SECTOR</th>
<th>DISEASE</th>
<th>SAVINGS</th>
<th><em>LEVEL</em></th>
<th>P</th>
<th>L</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>33</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.32745</td>
<td>0.25973</td>
<td>0.40322</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>35</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.34011</td>
<td>0.26901</td>
<td>0.41922</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>36</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.18418</td>
<td>0.12190</td>
<td>0.26855</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.51220</td>
<td>0.36329</td>
<td>0.65897</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>18</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.24109</td>
<td>0.18125</td>
<td>0.31314</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.28517</td>
<td>0.22445</td>
<td>0.35480</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.28517</td>
<td>0.22445</td>
<td>0.35480</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>31</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.31504</td>
<td>0.25007</td>
<td>0.38815</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>26</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.28517</td>
<td>0.22445</td>
<td>0.35480</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>37</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.35360</td>
<td>0.27793</td>
<td>0.43610</td>
</tr>
</tbody>
</table>
Predicted Values

Use the PREDPROBS = I option in order to obtain the predicted category (which is saved in the _INTO_ variable).

```plaintext
DM 'LOG; CLEAR; OUT; CLEAR;';

PROC LOGISTIC DATA = OB2 DESCENDING;
   MODEL DISEASE = AGE / CL;
   OUTPUT OUT = LOUT P=P L=L U=U PREDPROBS=I;
RUN;
```
Predicted Values

_FROM_ = The observed response category = The same value as the response variable.
Predicted Values

_into_ = The predicted response category.

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>S</th>
<th>I</th>
<th>A</th>
<th>F</th>
<th>T</th>
<th>L</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>O</td>
<td>A</td>
<td>S</td>
<td>T</td>
<td>N</td>
<td>O</td>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>b</td>
<td>P</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>D</td>
<td>E</td>
<td>S</td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>_</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>33</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>35</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>18</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>63</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>31</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>26</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>37</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Predicted Values

IP_1 = The Individual Probability of a response of 1.
Dear students, you have done well this semester by learning so many methods that are rarely taught to undergraduates!

For the final exam, the required materials would end here. From the next slide and on, those materials are not required.

As always, please read your little SAS book by Cody & Smith for logistic regression.
Scoring Observations in SAS

Obtaining predicted probabilities and/or predicted outcomes (categories) for new observations (i.e., scoring new observations) is done in logit modeling using the same procedure we used in scoring new observations in linear regression.
Scoring Observations in SAS

1. Create a new data set with the desired values of the x variables and the y variable set to missing.
2. Merge the new data set with the original data set.
3. Refit the final model using PROC LOGISTIC using the OUTPUT statement.
Classification Table & Rates

A Classification Table is used to summarize the results of the predictions and to ultimately evaluate the fitness of the model.

Obtain a classification table using PROC FREQ.

```
PROC FREQ DATA = LOUT;
   TABLES _FROM_*_INTO_;
RUN;
```
Classification Table & Rates

The observed (or actual) response is in rows and the predicted response is in columns.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0</strong></td>
<td>130</td>
<td>9</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>66.33</td>
<td>4.59</td>
<td>70.92</td>
</tr>
<tr>
<td></td>
<td>93.53</td>
<td>6.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>73.03</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24.49</td>
<td>4.59</td>
<td>29.08</td>
</tr>
<tr>
<td></td>
<td>84.21</td>
<td>15.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.97</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>178</td>
<td>18</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>90.82</td>
<td>9.18</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Correct classifications are summarized on the main diagonal.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percent</th>
<th>Row Pct</th>
<th>Col Pct</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>130</td>
<td>9</td>
<td>139</td>
<td>70.92</td>
</tr>
<tr>
<td></td>
<td>66.33</td>
<td>4.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>93.53</td>
<td>6.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>73.03</td>
<td>50.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>9</td>
<td>57</td>
<td>29.08</td>
</tr>
<tr>
<td></td>
<td>24.49</td>
<td>4.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84.21</td>
<td>15.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.97</td>
<td>50.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>178</td>
<td>18</td>
<td>196</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>90.82</td>
<td>9.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification Table & Rates

The total number of correct classifications (i.e., ‘hits’) is the sum of the main diagonal frequencies.

\[ O = 130 + 9 = 139 \]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0</strong></td>
<td>130</td>
<td>9</td>
<td>139</td>
</tr>
<tr>
<td>( \text{Row Pct} ) &amp; 66.33</td>
<td>4.59</td>
<td>70.92</td>
<td></td>
</tr>
<tr>
<td>&amp; 93.53</td>
<td>6.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&amp; 73.03</td>
<td>50.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td>48</td>
<td>9</td>
<td>57</td>
</tr>
<tr>
<td>( \text{Row Pct} ) &amp; 24.49</td>
<td>4.59</td>
<td>29.08</td>
<td></td>
</tr>
<tr>
<td>&amp; 84.21</td>
<td>15.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&amp; 26.97</td>
<td>50.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>178</td>
<td>18</td>
<td>196</td>
</tr>
<tr>
<td>( \text{Row Pct} ) &amp; 90.82</td>
<td>9.18</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>
The total-group hit rate is the ratio of O and N.  \( HR = \frac{139}{196} = .698 \)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Row Pct</th>
<th>Col Pct</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>130</td>
<td>9</td>
<td>139</td>
<td>70.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>66.33</td>
<td>4.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>93.53</td>
<td>6.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>73.03</td>
<td>50.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>9</td>
<td>57</td>
<td>29.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24.49</td>
<td>4.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84.21</td>
<td>15.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.97</td>
<td>50.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>178</td>
<td>18</td>
<td>196</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90.82</td>
<td>9.18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Individual group hit rates can also be calculated. These are essentially the row percents on the main diagonal.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>130</td>
<td>9</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>66.33</td>
<td>4.59</td>
<td>70.92</td>
</tr>
<tr>
<td></td>
<td>93.53</td>
<td>6.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>73.03</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>9</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>24.49</td>
<td>4.53</td>
<td>29.08</td>
</tr>
<tr>
<td></td>
<td>84.21</td>
<td>15.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.97</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>178</td>
<td>18</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>90.82</td>
<td>9.18</td>
<td>100.00</td>
</tr>
</tbody>
</table>
LDA vs. Logistic Regression

- LDA (Generative model)
  - Assumes Gaussian class-conditional densities and a common covariance
  - Model parameters are estimated by maximizing the full log likelihood, parameters for each class are estimated independently of other classes, $Kp+p(p+1)/2+(K-1)$ parameters
  - Makes use of marginal density information $Pr(X)$
  - Easier to train, low variance, more efficient if model is correct
  - Higher asymptotic error, but converges faster
LDA vs. Logistic Regression

- Logistic Regression (Discriminative model)
  - Assumes class-conditional densities are members of the (same) exponential family distribution
  - Model parameters are estimated by maximizing the conditional log likelihood, simultaneous consideration of all other classes, \((K-1)(\rho+1)\) parameters
  - Ignores marginal density information \(\Pr(X)\)
  - Harder to train, robust to uncertainty about the data generation process
  - Lower asymptotic error, but converges more slowly
# Generative vs. Discriminative Learning

(Rubinstein 97)

<table>
<thead>
<tr>
<th></th>
<th>Generative</th>
<th>Discriminative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example</strong></td>
<td>Linear Discriminant Analysis</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td><strong>Objective Functions</strong></td>
<td>Full log likelihood: $\sum_i \log p_o(x_i, y_i)$</td>
<td>Conditional log likelihood $\sum_i \log p_o(y_i</td>
</tr>
<tr>
<td><strong>Model Assumptions</strong></td>
<td>Class densities: $p(x</td>
<td>y = k)$ e.g. Gaussian in LDA</td>
</tr>
<tr>
<td><strong>Parameter Estimation</strong></td>
<td>“Easy” – One single sweep</td>
<td>“Hard” – iterative optimization</td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td>More efficient if model correct, borrows strength from $p(x)$</td>
<td>More flexible, robust because fewer assumptions</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Bias if model is incorrect</td>
<td>May also be biased. Ignores information in $p(x)$</td>
</tr>
</tbody>
</table>
Comparison between LDA and LOGREG

<table>
<thead>
<tr>
<th>True Distribution</th>
<th>LDA</th>
<th>LOGREG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly non-Gaussian</td>
<td>25.2/0.47</td>
<td>12.6/0.94</td>
</tr>
<tr>
<td>N/A</td>
<td>9.6/0.61</td>
<td>4.1/0.17</td>
</tr>
<tr>
<td>Gaussian</td>
<td>7.6/0.12</td>
<td>8.1/0.27</td>
</tr>
</tbody>
</table>

(Rubinstein 97)
Link between LDA, QDA, and Logistic regression

Suppose we classify by computing a posterior probability. The posterior was calculated by modeling the likelihood and prior for each class.

- In LDA & QDA, we compute the posterior, assuming that each class distribution was Gaussian with equal (LDA) or unequal (QDA) variances.

- In logistic regression, we directly model the posterior as a function of the variable $x$.
  \[
  \hat{P}(l|x) = g(x)
  \]

- In practice, when there are $k$ classes to classify, we model:
  \[
  \frac{P(1|x)}{P(k|x)} = g_1(x)
  \]
Classification by maximizing the posterior distribution

In this example we assume that the two distributions for the classes have **equal variance**. Suppose we want to classify a person as male or female based on height.

What we have: \( p(x \mid y = 0) \) and \( p(x \mid y = 1) \) and \( P(y = 1) = q \)

What we want: \( P(y = 1 \mid x) \)

Height is normally distributed in the population of men and in the population of women, with different means, and similar variances. Let \( y \) be an indicator variable for being a female. Then the conditional distribution of \( x \) (the height becomes):

\[
p(x \mid y = 1) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2} \left( x - \mu_f \right)^2 \right)
\]

\[
p(x \mid y = 0) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2} \left( x - \mu_m \right)^2 \right)
\]
Posterior probability for classification when we have two classes of equal variance:

\[ P(y = 1 | x) = \frac{P(y = 1) p(x | y = 1)}{P(y = 1) p(x | y = 1) + P(y = 0) p(x | y = 0)} \]

\[ = \frac{q \exp\left( -\frac{1}{2\sigma^2} (x - \mu_f)^2 \right)}{q \exp\left( -\frac{1}{2\sigma^2} (x - \mu_f)^2 \right) + (1-q) \exp\left( -\frac{1}{2\sigma^2} (x - \mu_m)^2 \right)} \]

\[ = \frac{1}{1 + \frac{(1-q) \exp\left( -\frac{1}{2\sigma^2} (x - \mu_m)^2 \right)}{q \exp\left( -\frac{1}{2\sigma^2} (x - \mu_f)^2 \right)}} \]

\[ = \frac{1}{1 + \exp\left( \log\left( \frac{1-q}{q} \right) - \frac{1}{2\sigma^2} \left( (x - \mu_m)^2 - (x - \mu_f)^2 \right) \right)} \]

\[ = \frac{1}{1 + \exp\left( \log\left( \frac{1-q}{q} \right) - \frac{1}{2\sigma^2} \left( \mu_m^2 - \mu_f^2 \right) + \frac{\mu_m - \mu_f}{\sigma^2} x \right)} \]
Computing the probability that the subject is female, given that we observed height $x$.

$$P(y = 1 | x) = \frac{1}{1 + \exp \left( \log \left( \frac{1-q}{q} \right) - \frac{1}{2\sigma^2} (\mu_m^2 - \mu_f^2) + \frac{(\mu_m - \mu_f)}{\sigma^2} x \right) }$$

In the denominator, $x$ appears linearly inside the exponential.

So if we assume that the class membership densities $p(x | y)$ are normal with equal variance, then the posterior probability will be a logistic function.
Logistic regression with assumption of equal variance among density of classes implies a linear decision boundary.

\[
P\left(y^{(i)} = 1 \mid x^{(i)}\right) = \frac{1}{1 + \exp\left(a_0 - a^T x^{(i)}\right)}
\]

\[
P\left(y^{(i)} = 0 \mid x^{(i)}\right) = 1 - \frac{1}{1 + \exp\left(a_0 - a^T x^{(i)}\right)} = \frac{\exp\left(a_0 - a^T x^{(i)}\right)}{1 + \exp\left(a_0 - a^T x^{(i)}\right)}
\]

\[
y^{(i)} = 1 \text{ if } \log \frac{P\left(y^{(i)} = 1 \mid x^{(i)}\right)}{P\left(y^{(i)} = 0 \mid x^{(i)}\right)} > 0
\]

\[
\log \frac{P\left(y^{(i)} = 1 \mid x^{(i)}\right)}{P\left(y^{(i)} = 0 \mid x^{(i)}\right)} = \log \frac{1}{\exp\left(a_0 - a^T x^{(i)}\right)} = -a_0 + a^T x^{(i)}
\]
Modeling the posterior when the densities have **unequal variance**

\[
P(y = 1 | x) = \frac{P(y = 1) p(x | y = 1)}{P(y = 1) p(x | y = 1) + P(y = 0) p(x | y = 0)}
\]

\[
= \frac{q \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma_1^2} (x - \bar{x}_1)^2 \right)}{1 + (1-q) \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma_2^2} (x - \bar{x}_2)^2 \right) + \frac{q}{\sigma_1 \sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma_1^2} (x - \bar{x}_1)^2 \right)}
\]

\[
= \frac{1}{1 + \exp \left( \log \left( \frac{1-q}{q} \right) + \log \left( \frac{\sigma_1}{\sigma_2} \right) - \frac{1}{2\sigma_2^2} (x - \bar{x}_2)^2 + \frac{1}{2\sigma_1^2} (x - \bar{x}_1)^2 \right)}
\]

\[
= \frac{1}{1 + \exp \left( w_0 + w_1 x + w_2 x^2 \right)}
\]
Related sites/papers:

- ftp://gis.msl.mt.gov/Maxell/Models/Predictive_Modeling_for_DSS_Lincoln_NE_121510/Modeling_Literature/Efron_Efficiency%20of%20LR%20versus%20DFA.pdf
- http://math.arizona.edu/~hzhang/math574m/Read/LogitOrLDA.pdf
- https://onlinecourses.science.psu.edu/stat857/node/184
Related sites/papers:

Related sites/papers:

Acknowledgement:

- We also thank colleagues who posted their lecture notes on the internet!