# 1、本项目旨在用python实现传统逻辑回归方法建模的全过程，包括两大环节：

1、原始数据到建模数据。这一环节包括：

1）数据导入

2）数据质量检查和处理（缺失值等）

3）构建表现变量和预测变量

4）模型分组

5）抽样

2、基于建模数据进行建模。这一环节包括以下内容：

1）细分栏，输出细分的变量格式

2）将数据集基于变量格式输出woe信息表

3）根据woe信息表把建模数据转换为woe数据

4）通过逐步判别分析法（暂定）挑选变量

5）粗分栏，并转换woe，输出woe表

6）逻辑回归建模，实现stepwise方法

7）计算模型性能指标：KS、ROC、Gini、Divergence，并绘图。

# 2、stepdisc过程的实现

**Wilks'λ值**的统计量

Wilks'lambda是组内平方和与总平方和之比。当所有观测的组均值相等时，Wilks'lambda值为1；当组内变异与总变异相比小时，Wilks'lambda值接近于0。因此，Wilks'lambda值大，表示各个组的均值基本相等；Wilks'lambda小表示组间有差异。在判别分析中，只有组均值不等时，判别分析才有意义。  
另外，在步进法中wilks'lambda是一种用于逐步判别分析的变量选择方法，是基于变量能够在多大程度上降低wilks'lambda值来选择要输入到方程中的变量，每一步中局势能够使总体wilks'lambda值最小的变量。Wilk’s lambda 值在0-1之间，值越小贡献越大。

1、首先计算wilks lambda简称WL

WL是两个矩阵的**行列式**的比值，分子是组内变量之间的协方差，然后几组协方差矩阵加起来。分母是总体的协方差矩阵。比如，变量的列矩阵为 Xmn，m个观测，n个变量，则协方差矩阵是n\*n的。然后分组，各自算各自的协方差矩阵 。然后算行列式值。

2、WL的值越小，说明变量之间越不相关，越好。

3、如果已经有l个变量选中了，则要算WL(l+1)/WL(l)这个统计量，也是越小越好，然后根据WL(l+1)/WL(l)-1服从F(z-1,n-l-z)分布，来检验给定前l个变量的条件下，再增加这个变量的条件均值是否相等，即是否对区分总体提供附加信息。其中，z是组别数量，l是已经选中的变量数，n是样本量

ilks Lambda is the proportion of total variance in the Discriminant Scores that is not explained by the difference between groups . It is equal to 1-RSQ, i.e. is equal to ( 1 - (canonical Correlation)^2 )

## 什么是威尔克斯的Lambda？

Wilks的lamdba（Λ）是一项检验统计量，在[MANOVA](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/" \l "MANOVA)，判别分析和其他多元过程的结果中进行了报告。其他类似的测试统计数据包括**Pillai的跟踪准则**和**Roy ger准则。**

* 在[**MANOVA中**](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/#MANOVA)，Λ检验[因变量](https://www.statisticshowto.com/dependent-variable-definition/" \t "_blank)的特定组合的组均值之间是否存在差异。它类似于[F-检验统计量](https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/)的[方差分析](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/#ANOVA)。Lambda是因变量的方差百分比的量度，而因变量的水平差异无法解释。零值表示没有任何自变量无法解释的方差（理想情况）。换句话说，统计量越接近零，所讨论的变量对模型的贡献就越大。当Wilk的lambda接近于零时，您将拒绝原假设，尽管这应与较小的p值结合使用。
* 在**判别分析中**，Wilk的lambda测试了每个级别的[自变量](https://www.statisticshowto.com/independent-variable-definition/)对模型的贡献程度。比例尺范围从0到1，其中0表示完全区分，而1表示没有区分。通过将每个自变量放入模型中然后将其取出进行测试-生成Λ统计量。Λ的变化的显着性通过[F检验测量](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/f-test/" \t "_blank)；如果F值大于[临界值](https://www.statisticshowto.com/probability-and-statistics/find-critical-values/)，则变量将保留在模型中。通常使用Minitab，R或SPSS之类的软件执行此逐步过程。以下SPSS输出显示使用此过程保留了哪些变量（从十几个或更多列表中）。

分母中的1 –λ是因变量[的方差比例，](https://www.statisticshowto.com/proportion-of-variance/)由模型效应解释。在解释结果时应格外小心，因为该统计数据容易产生[偏差](https://www.statisticshowto.com/what-is-bias/)，尤其是对于小样本而言。

## 输出组件

Wilks的lambda输出包含几个部分，包括：

* “ Sig”或显着性（[p值](https://www.statisticshowto.com/p-value/" \t "_blank)）。如果该值很小，则（即低于0.05）[拒绝原假设](https://www.statisticshowto.com/support-or-reject-null-hypothesis/)。
* 输出中的“值”列：Wilk Lambda的值。
* “统计”是与列出的自由度相关的F统计。将以APA格式报告为F（df1，df2）=值。例如，如果您有一个自由度为1和2的f值36.612，则将其报告为F（1,2）= 36.612。

**参考文献**

Nath, R. and Pavur, R. (1985) A new statistic in the one way multivariate analysis of variance, Computational Statistics and Data Analysis, 2, 297–315.  
Todorov, V. and Filzmoser, P. (2007) Robust statistic for the one-way MANOVA, submitted to the Journal of Environmetrics.

逐步选择：开始时如同向前选择一样，模型中没有变量，每一步都被检查。如果在Wilks的 准则下统计量对模型的判别能力贡献最小的变量达不到留在模型中的标准，它就被剔除。否则，不在模型中对模型的判别能力贡献最大的变量被选入模型。当模型中的所有变量都达到留在模型中的标准而没有其他变量能达到进入模型的标准，逐步选择过程停止。

**3、选择准则选项**  
  
**1）SLENTRY= |SLE=** ：在向前选择方法中，指定选入变量的显著性水平，  
**2）SLSTAY= |SLS=** ：在向后剔除方法中，指定保留变量的显著性水平， 。缺省值为0.15。  
**3）PR2ENTRY＝p|PR2E＝p**：在向前选择方式中，指定选入变量的偏 ， 。  
**4）PR2STAV＝p|PR2S＝p**：在向后剔除方式中，指定保留变量的偏 ， 。  
  
**4、选择过程选项**  
  
**1）INCLUDE＝n**：要求VAR语句中前 个变量包含在每一个模型中。缺省值为0。  
**2）MAXSTEP＝n**：指定最多步数。缺省值是VAR语句中变量数的两倍。  
**3） START＝n**： 指定VAR语句中前n个变量被用来开始选择过程。当选用METHOD＝FORWARD或METHOD＝STEPWISE时，缺省值为0；当选用METHOD＝BACKWARD时，缺省为VAR语句中的变量数。  
**4）STOP＝n**： 指定最终模型中的变量数。一旦发现包含n个变量的模型，STFPDISC过程便停止选择过程。该选项只有当选用METHOD＝FORWARD或METHOD＝BACKWARD时才选用METHOD＝FORWARD时，缺省值为VAR语句中变量的个数；选用METHOD＝BACKWARD时，缺省值为0。

**5、奇异性选择**  
  
**SINQULAR=** ：指定选入变量的奇异标准， 。STEPDISC过程当一个变量与已在模型中的变量的平方多重相关超过 时，拒绝该变量进入模型。当模型中不止一个变量时，如果一个变量进入模型回使得已在模型中的任何一个变量与该变量及其已在模型中变量的平方多重相关超过 ，STEPDISC也拒绝该变量，缺省值为 。

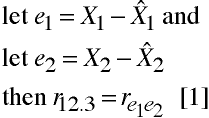
协方差分析中用f检验

squared partial correlation 平方偏相关

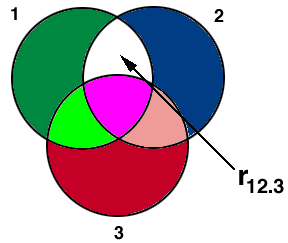
Partial Correlation

r12.3 is the correlation between variables 1 and 2 with variable 3 removed from both variables. To illustrate this, run separate regressions using X3 as the independent variable and X1 and X2 as dependent variables. Next, compute residuals for regression...

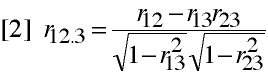
 **r12.3** is the correlation between variables 1 and 2 with variable 3 removed from both variables. To illustrate this, run separate regressions using **X3** as the independent variable and **X1** and **X2** as dependent variables. Next, compute residuals for regression...



**Venn Diagram of Partial Correlation**

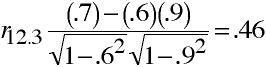


**More Partial Correlation**

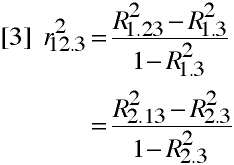


**Example**

 **Let r12 = 0.7;   r13 = 0.6;   r23 = 0.9.**



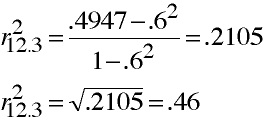
**Using Multiple Correlations**



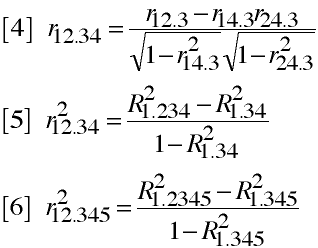
 Thus, squared partial correlations represent the ratio of incremental variance to the residual variance.

**Example**

 **Let R21.23 = .4947 and R21.3 = .62 = .36** (from the above example)



**Higher Order Partial Correlation**



and so on...

**Partial R-squares**

The partial R-square (or coefficient of partial determination) measures the

marginal contribution of one explanatory variable when all others are

already included in the model. For example, in the above example only 7.7%

of the variation in mpg is reduced by adding weight to the model when length

is already in the model. However, such interpretation is not valid, for

example, if your explanatory variables are collinear

方差分析(Analysis of Variance，简称ANOVA)

**partial** R2R2 (or it is also called the **coefficient of partial determination**).

# 3、logistic回归求解算法

## 3.1 sklearn包里面的logisticRegression

包括

1）sklearn.linear\_model.LogisticRegressionCV

2）sklearn.linear\_model.LogisticRegression

算法：“liblinear”, “newton-cg”, “lbfgs”, “sag” and “saga”:

1) The “lbfgs”是默认算法。

* For small datasets, ‘liblinear’ is a good choice, whereas ‘sag’ and ‘saga’ are faster for large ones.
* For multiclass problems, only ‘newton-cg’, ‘sag’, ‘saga’ and ‘lbfgs’ handle multinomial loss; ‘liblinear’ is limited to one-versus-rest schemes.
* ‘newton-cg’, ‘lbfgs’, ‘sag’ and ‘saga’ handle L2 or no penalty
* ‘liblinear’ and ‘saga’ also handle L1 penalty
* ‘saga’ also supports ‘elasticnet’ penalty
* ‘liblinear’ does not support setting penalty='none'
* The solver “liblinear” 使用坐标下降法- coordinate descent (CD) algorithm。
* The “lbfgs” solver is used by default for its robustness. For large datasets the “saga” solver is usually faster. For large dataset, you may also consider using SGDClassifier with ‘log’ loss, which might be even faster but requires more tuning.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Solvers** | | | | |
| **Penalties** | **‘liblinear’** | **‘lbfgs’** | **‘newton-cg’** | **‘sag’** | **‘saga’** |
| Multinomial + L2 penalty | no | yes | yes | yes | yes |
| OVR + L2 penalty | yes | yes | yes | yes | yes |
| Multinomial + L1 penalty | no | no | no | no | yes |
| OVR + L1 penalty | yes | no | no | no | yes |
| Elastic-Net | no | no | no | no | yes |
| No penalty (‘none’) | no | yes | yes | yes | yes |
| **Behaviors** |  | | | | |
| Penalize the intercept (bad) | yes | no | no | no | no |
| Faster for large datasets | no | no | no | yes | yes |
| Robust to unscaled datasets | yes | yes | yes | no | no |

## 3.2 SAS使用的proc logistic

假设是解释变量的向量，并且是要建模的响应概率。线性逻辑模型具有形式

|  |  |  |
| --- | --- | --- |
|  |  |  |

默认情况下，LOGISTIC过程对较低响应级别的概率进行建模。

LOGISTIC程序通过最大似然法拟合离散响应数据的线性逻辑回归模型。它还可以对二进制响应数据执行条件逻辑回归，对二进制和名义响应数据执行精确的条件逻辑回归。最大似然估计是使用Fisher评分算法或Newton-Raphson算法执行的，您可以执行Firth（[1993](https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/statug_logistic_sect067.htm#firt_d_93)）和Heinze and Schemper（[2002](https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/statug_logistic_sect067.htm#hein_g_2002)）讨论的减少偏差的惩罚似然优化。。您可以指定参数估计值的起始值。logistic回归模型中的logit链接函数可以用probit函数，互补log-log函数或广义logit函数代替。

The LOGISTIC procedure fits linear logistic regression models for discrete response data by the method of maximum likelihood. It can also perform conditional logistic regression for binary response data and exact logistic regression for binary and nominal response data. The maximum likelihood estimation is carried out with either the **Fisher scoring algorithm** or the **Newton-Raphson algorithm**, and you can perform the bias-reducing penalized likelihood optimization as discussed by Firth ([1993](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_references.htm&docsetVersion=14.3&locale=en#statug_logisticfirt_d93)) and Heinze and Schemper ([2002](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_references.htm&docsetVersion=14.3&locale=en#statug_logistichein_g2002)). You can specify starting values for the parameter estimates.

* [迭代加权最小二乘算法（Fisher评分）](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_details05.htm&docsetVersion=14.3&locale=en#statug.logistic.logisticirlsa)
* [牛顿-拉夫森算法](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_details05.htm&docsetVersion=14.3&locale=en#statug_logistic007360)

默认的最大似然算法是Fisher评分方法，等效于通过迭代重新加权的最小二乘拟合。替代算法是Newton-Raphson方法。对于广义logit模型，相邻类别的logit模型以及指定[EQUALSLOPES](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticequal)或[UNEQUALSLOPES的模型](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticunequal)选项，仅牛顿-拉夫森技术可用。两种算法产生相同的参数估计值。但是，参数估计量的估计协方差矩阵可能会略有不同，因为Fisher评分基于预期的信息矩阵，而Newton-Raphson方法则基于观测的信息矩阵。对于二进制logit模型，观测和预期信息矩阵相同，因此两种算法的估计协方差矩阵相同。您可以指定[TECHNIQUE =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logistictechnique)选项来选择拟合算法，还可以指定[FIRTH](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticmfirth)选项来执行减少偏差的惩罚性最大似然拟合。

3.3 statsmodel里面使用的logit regression

**效果选择方法**

通过在[MODEL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)语句中指定[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticselection)选项，可以使用五种效果选择方法。最简单的方法（默认设置）是[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticselection) NONE，对于该方法，PROC LOGISTIC适合[MODEL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)语句中指定的完整模型。其他四种方法是：FORWARD用于正向选择，BACKWARD用于反向消除，STEPWISE用于逐步选择，SCORE用于最佳子集选择。除非指定了[NOINT](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticnoint)选项，否则截取参数将强制保留在模型中。

当[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en#statug.logistic.logisticselection) FORWARD时，PROC LOGISTIC首先估算 强制进入模型的效果参数。这些效果是截距和第一*Ñ*在解释性效果[MODEL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)语句，其中*Ñ*是由指定数量[START =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstart)或[INCLUDE =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticinclude)在选项[MODEL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)声明（*Ñ*是零通过默认值）。接下来，该过程将为模型中未包含的每个效应计算**得分卡方统计量（ the score chi-square statistic）**，并检查这些统计量中最大的一个。如果[SLENTRY =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en#statug.logistic.logisticslentry)级别，相应的效果将添加到模型中。一旦在模型中输入了效果，就永远不会将其从模型中删除。重复该过程，直到所有剩余效果都未达到指定的输入级别或达到[STOP =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstop)值为止。有关更多信息，请参见“ [分数统计和测试](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_details14.htm&docsetVersion=14.3&locale=en) ”部分。

当[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en#statug.logistic.logisticselection) BACKWARD时，参数为完整除非指定了[START =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstart)选项，否则将估算[MODEL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)语句中 指定的[模型](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)。在那种情况下，仅估计截距的参数和[MODEL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en)语句中的前*n个*说明性效果，其中*n*是START =选项指定的数字。检验了Wald测试各个参数的结果。删除了不符合留在模型中的[SLSTAY =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticslstay)级别的最小显着影响。从模型中删除效果后，该效果将保持排除状态。重复该过程，直到模型中没有其他效果达到指定的消除水平或直到[STOP =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstop)达到价值。向后选择通常不如向前或逐步选择成功，因为第一步中的完整模型拟合是最有可能导致响应值完全或准完全分离的模型，如[最大似然估计的存在](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_details10.htm&docsetVersion=14.3&locale=en)部分中所述。有关更多信息，请参见[测试关于回归系数的线性假设](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_details36.htm&docsetVersion=14.3&locale=en)。

所述[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticselection) STEPWISE选项是类似于 [SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en#statug.logistic.logisticselection) FORWARD选项，只是模型中已经存在的效果不一定会保留。将效果输入到模型中或从模型中删除效果，使得每个前向选择步骤之后都可以执行一个或多个后向消除步骤。如果无法进一步增加模型效果，或者当前模型与以前访问的模型相同，则逐步选择过程将终止。

对于[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticselection) SCORE，PROC LOGISTIC使用 Furnival和Wilson的分支定界算法（[1974](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_references.htm&docsetVersion=14.3&locale=en" \l "statug_logisticfurn_g74)）适用于从1，2，3效果模型等直到所有可能的模型尺寸，找到具有最高得分卡方统计量的指定数量的模型。包含所有解释性影响的单个模型。每种型号尺寸显示的型号数量由[BEST =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticmbest)选项控制。您可以使用[START =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstart)选项强加最小模型尺寸，并且可以使用[STOP =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstop)选项强加最大模型尺寸。例如，对于[BEST =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticmbest) 3，[START =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstart) 2和[STOP =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstop)在图5中，SCORE选择方法显示包含2、3、4和5个效果的三个最佳模型（即，具有最高得分卡方统计的三个模型）。该[SELECTION =](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticselection) SCORE选项不适用于包含类变量或模型时，你还可以指定LINK = GLOGIT。有关更多信息，请参见“ [分数统计和测试](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_details14.htm&docsetVersion=14.3&locale=en) ”部分。

当选项[FAST](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en#statug.logistic.logisticfast)，[SEQUENTIAL](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticsequential)和[STOPRES](https://documentation.sas.com/?docsetId=statug&docsetTarget=statug_logistic_syntax22.htm&docsetVersion=14.3&locale=en" \l "statug.logistic.logisticstopres)与FORWARD，BACKWARD或STEPWISE选择方法一起使用时，它们可以更改用于输入模型效果或从模型中删除效果的默认标准。

### Effect-Selection Methods

Five effect-selection methods are available by specifying the [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) option in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement. The simplest method (and the default) is [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) NONE, for which PROC LOGISTIC fits the complete model as specified in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement. The other four methods are FORWARD for forward selection, BACKWARD for backward elimination, STEPWISE for stepwise selection, and SCORE for best subsets selection. Intercept parameters are forced to stay in the model unless the [NOINT](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticnoint) option is specified.

When [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) FORWARD, PROC LOGISTIC first estimates parameters for effects forced into the model. These effects are the intercepts and the first n explanatory effects in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement, where n is the number specified by the [START=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstart) or [INCLUDE=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticinclude) option in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement (n is zero by default). Next, the procedure computes the score chi-square statistic for each effect not in the model and examines the largest of these statistics. If it is significant at the [SLENTRY=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticslentry) level, the corresponding effect is added to the model. Once an effect is entered in the model, it is never removed from the model. The process is repeated until none of the remaining effects meet the specified level for entry or until the [STOP=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstop) value is reached.

When [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) BACKWARD, parameters for the complete model as specified in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement are estimated unless the [START=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstart) option is specified. In that case, only the parameters for the intercepts and the first n explanatory effects in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement are estimated, where n is the number specified by the START= option. Results of the Wald test for individual parameters are examined. The least significant effect that does not meet the [SLSTAY=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticslstay) level for staying in the model is removed. Once an effect is removed from the model, it remains excluded. The process is repeated until no other effect in the model meets the specified level for removal or until the [STOP=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstop) value is reached. Backward selection is often less successful than forward or stepwise selection because the full model fit in the first step is the model most likely to result in a complete or quasi-complete separation of response values as described in the section [Existence of Maximum Likelihood Estimates](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details10.htm).

The [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) STEPWISE option is similar to the [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm" \l "statug.logistic.logisticselection) FORWARD option except that effects already in the model do not necessarily remain. Effects are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further effect can be added to the model or if the current model is identical to a previously visited model.

For [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) SCORE, PROC LOGISTIC uses the branch-and-bound algorithm of Furnival and Wilson ([1974](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_references.htm#statug_logisticfurn_g74)) adapted to find a specified number of models with the highest likelihood score (chi-square) statistic for all possible model sizes, from 1, 2, 3 effect models, and so on, up to the single model containing all of the explanatory effects. The number of models displayed for each model size is controlled by the [BEST=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmbest) option. You can use the [START=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstart) option to impose a minimum model size, and you can use the [STOP=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstop) option to impose a maximum model size. For instance, with [BEST=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmbest) 3, [START=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstart) 2, and [STOP=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstop) 5, the SCORE selection method displays the best three models (that is, the three models with the highest score chi-squares) containing 2, 3, 4, and 5 effects. The [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) SCORE option is not available for models with CLASS variables.

The options [FAST](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticfast) , [SEQUENTIAL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticsequential) , and [STOPRES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticstopres) can alter the default criteria for entering or removing effects from the model when they are used with the FORWARD, BACKWARD, or STEPWISE selection method.

SLENTRY=value   
SLE=value

specifies the significance level of the score chi-square for entering an effect into the model in the FORWARD or STEPWISE method. Values of the SLENTRY= option should be between 0 and 1, inclusive. By default, SLENTRY=0.05. The SLENTRY= option has no effect when [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) NONE, [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) BACKWARD, or [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) SCORE.

SLSTAY=value   
SLS=value

specifies the significance level of the Wald chi-square for an effect to stay in the model in a backward elimination step. Values of the SLSTAY= option should be between 0 and 1, inclusive. By default, SLSTAY=0.05. The SLSTAY= option has no effect when [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) NONE, [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) FORWARD, or [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) SCORE.

Score检验和wald检验都有检验一个变量=0和多个变量=0的统计量公式，分布~卡方（1）和~卡方（q）q个变量。

似然比（LR）检验、 沃德检验、拉格朗日乘数检验（LM/即SCORE检验）之间的不等式，w>=LR>=LM。三者近似相等，但有限样本中数值结果有所不同。

**问题：**

Sas所用的方法是用wald检验 score检验进行单个参数检验还是联合显著性检验？？

**1、选变量用的是the score chi-square即得分卡方统计量**

得分向量也称为斜率向量或梯度向量，其实质是对数似然函数LnL（x，θ）对参数θ（k维向量）一阶偏导数所构成的k维向量。LM检验的基本原理是：如果约束h（θ）=0是有效的，则最大化拉格朗日函数LnL（R）（x，θ）所得到的有约束的参数估计量θ'应该位于最大化原始样本似然函数LnL（x，θ）的参数估计值附近。因此对数似然函数的斜率（也称得分向量）在θ'处应该趋近于0。于是，LM检验就是在有约束估计量θ'处，通过检验得分向量是否趋近于0来检验约束是否有效，这种检验方法也因此被称为得分检验。

在零假设[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image074%5b3%5d.gif)：[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image076%5b1%5d.gif)＝0下，设参数的估计值为[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image122_2.gif)，即对应的[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image076%5b2%5d.gif)＝0。计算Score统计量的公式为

[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image124_2.gif)　　　　　　　　　　（2.6）

上式中，[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image126_2.gif)表示在[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image076%5b3%5d.gif)＝0下的对数似然函数（1.9）的一价偏导数值，而[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image128_2.gif)表示在[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image076%5b4%5d.gif)＝0下的对数似然函数（1.9）的二价偏导数值。Score统计量服从自由度等于１的[](http://images.cnblogs.com/cnblogs_com/zgw21cn/WindowsLiveWriter/d5cfb49e1453_1013A/clip_image084%5b2%5d.gif)分布（应该是chi方分布）。

**2、剔变量用的是the Wald chi-square即沃德卡方统计量。**

Wald卡方检验是用来检验模型的参数的。大多数情况下和T检验很接近，统计量类似T值的平方。

两种检验方法Wald-score test 针对以下假设检验问题：

0: = 0, : ≠

其根本的检验逻辑是一样的。

根本逻辑： 根据样本得出参数 β的估计 ̂β^，然后通过比较 ̂β^与 β的“距离”；如果距离近，则不拒绝原假设；距离远即可拒绝原假设。 根本逻辑如上，不过两种方法对于“距离”的理解和处理有所不同。

**Wald方法中**，以大样本情况为例，“距离”的表示是我们所熟知的

=( ̂− 0)/ ( ̂)

当 = 时，z近似服从N(0,1). z是在某种规则标准化后的距离，越接近0，则表示距离越近，越不拒绝原假设。

**Score test方法中，**将 ̂与 0的差异表示为 ( ̂)和 ( 0)的差异，其中u( )为函数。在Score test里，u( )即为 ( )/ ，其中L为对数似然函数。

同理，也需要对这个差异进行标准化处理，需要除以− [ ^2 ( )/ ^2]。所以Score test方法里的“距离”最终表示为



注意：在极大似然估计的情况下， ( ̂)=0，上式可进一步化简。

3、第一步是入截距。怎么计算的。

4、模型收敛指什么，即**Model Convergence Status** Convergence criterion (GCONV=1E-8) satisfied. 指的是梯度下降等模型算法中的收敛条件，也就是提升度小到什么程度的时候，不再优化计算。

### Convergence Criteria

Four convergence criteria are available: [ABSFCONV=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmabsfconv) , [FCONV=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmfconv) , [GCONV=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmgconv) , and [XCONV=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticxconv) . If you specify more than one convergence criterion, the optimization is terminated as soon as one of the criteria is satisfied. If none of the criteria is specified, the default is GCONV=1E–8.

If you specify a [STRATA](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax34.htm) statement or the [EQUALSLOPES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticequal) or [UNEQUALSLOPES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticunequal) options in the MODEL statement, all unspecified (or nondefault) criteria are also compared to zero. For example, specifying only the criterion XCONV=1E–8 but attaining FCONV=0 terminates the optimization even if the XCONV= criterion is not satisfied, because the log likelihood has reached its maximum. More convergence criteria are also available; for more information, see the section [NLOPTIONS Statement](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax25.htm).

GCONV=value

specifies the relative gradient convergence criterion. Convergence requires that the normalized prediction function reduction is small,

\[  \frac{\mb{g}_ i^{\prime } \bI _ i^{-1} \mb{g}_ i}{|l_ i| + {\mbox{1E--6}}} < \mi{value}  \]

where $l_ i$is the value of the log-likelihood function, $\mb{g}_ i$is the gradient vector, and $\bI _ i$is the negative (expected) Hessian matrix, all at iteration i. This is the default convergence criterion, and the default value is 1E–8. See the section [Convergence Criteria](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details09.htm) for more information.

### Iterative Algorithms for Model Fitting

**Subsections:**

* [**Iteratively Reweighted Least Squares Algorithm (Fisher Scoring)**](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details05.htm#statug.logistic.logisticirlsa)
* [**Newton-Raphson Algorithm**](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details05.htm#statug_logistic007078)
* [**Firth’s Bias-Reducing Penalized Likelihood**](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details05.htm#statug.logistic.logisticdfirth)

This section describes the two iterative maximum likelihood algorithms that are available in PROC LOGISTIC for fitting an unconditional logistic regression. For information about available optimization techniques for conditional logistic regression and models that specify the [EQUALSLOPES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticequal) or [UNEQUALSLOPES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticunequal) options, see the section [NLOPTIONS Statement](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax25.htm). Exact logistic regression uses a special algorithm, which is described in the section [Exact Conditional Logistic Regression](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details53.htm).

The default maximum likelihood algorithm is the Fisher scoring method, which is equivalent to fitting by iteratively reweighted least squares. The alternative algorithm is the Newton-Raphson method. For generalized logit models and models that specify the [EQUALSLOPES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticequal) or [UNEQUALSLOPES](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticunequal) options, only the Newton-Raphson technique is available. Both algorithms produce the same parameter estimates. However, the estimated covariance matrix of the parameter estimators can differ slightly because Fisher scoring is based on the expected information matrix whereas the Newton-Raphson method is based on the observed information matrix. For a binary logit model, the observed and expected information matrices are identical, resulting in identical estimated covariance matrices for both algorithms. You can specify the [TECHNIQUE=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logistictechnique) option to select a fitting algorithm, and you can specify the [FIRTH](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmfirth) option to perform a bias-reducing penalized maximum likelihood fit.

#### Iteratively Reweighted Least Squares Algorithm (Fisher Scoring)

Consider the multinomial variable $\bZ _ j=(Z_{1j},\ldots ,Z_{k+1,j})’$such that

\begin{eqnarray*}  Z_{ij}= &  \left\{  \begin{array}{ll} 1 &  \mbox{if } Y_ j=i \\ 0 &  \mbox{otherwise} \end{array} \right. \\ \end{eqnarray*}

With $\pi _{ij}$denoting the probability that the *j*th observation has response value *i*, the expected value of $\bZ _ j$is $\bpi _ j=(\pi _{1j},\ldots ,\pi _{k+1,j})^\prime $where $\pi _{k+1,j}=1-\sum _{i=1}^ k\pi _{ij}$. The covariance matrix of $\bZ _ j$is $\bV _ j$, which is the covariance matrix of a multinomial random variable for one trial with parameter vector $\bpi _ j$. Let $\bbeta $be the vector of regression parameters; in other words, $\bbeta = (\alpha _{1}, \ldots , \alpha _{k}, \beta _{1},\ldots ,\beta _{s})’$. Let $\bD _ j$be the matrix of partial derivatives of $\bpi _ j$with respect to $\bbeta $. The estimating equation for the regression parameters is

\[  \sum _ j\bD ’_ j\bW _ j(\bZ _ j-\bpi _ j)=\bm {0}  \]

where $ {\bW }_ j=w_ j f_ j {\bV }_ j^-$, $w_ j$and $f_ j$are the weight and frequency of the jth observation, and ${\bV }_ j^-$is a generalized inverse of ${\bV }_ j$. PROC LOGISTIC chooses ${\bV }_ j^-$as the inverse of the diagonal matrix with $\bpi _ j$as the diagonal.

With a starting value of $\bbeta ^{(0)}$, the maximum likelihood estimate of $\bbeta $is obtained iteratively as

\[  \bbeta ^{(m+1)}=\bbeta ^{(m)}+(\sum _ j \bD ’_ j\bW _ j\bD _ j)^{-1} \sum _ j\bD ’_ j\bW _ j(\bZ _ j-\bpi _ j)  \]

where $\bD _ j$, $\bW _ j$, and $\bpi _ j$are evaluated at $\bbeta ^{(m)}$. The expression after the plus sign is the step size. If the likelihood evaluated at $\bbeta ^{(m+1)}$is less than that evaluated at $\bbeta ^{(m)}$, then $\bbeta ^{(m+1)}$is recomputed by step-halving or ridging as determined by the value of the [RIDGING=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticridging) option. The iterative scheme continues until convergence is obtained—that is, until $\bbeta ^{(m+1)}$is sufficiently close to $\bbeta ^{(m)}$. Then the maximum likelihood estimate of $\bbeta $is ${\widehat{\bbeta }}=\bbeta ^{(m+1)}$.

The covariance matrix of ${\widehat{\bbeta }}$is estimated by

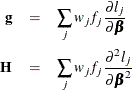
\[  \widehat{\mbox{Cov}}({\widehat{\bbeta }})=(\sum _ j \widehat{\bD }’_ j\widehat{\bW }_ j\widehat{\bD }_ j)^{-1}=\widehat{\bI }^{-1}  \]

where $\widehat{\bD }_ j$and $\widehat{\bW }_ j$are, respectively, $\bD _ j$and $\bW _ j$evaluated at ${\widehat{\bbeta }}$. $\widehat{\bI }$is the information matrix, or the negative expected Hessian matrix, evaluated at ${\widehat{\bbeta }}$.

By default, starting values are zero for the slope parameters, and for the intercept parameters, starting values are the observed cumulative logits (that is, logits of the observed cumulative proportions of response). Alternatively, the starting values can be specified with the [INEST=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax01.htm#statug.logistic.logisticinest) option.

#### Newton-Raphson Algorithm

For cumulative models, let the parameter vector be $\bbeta =(\alpha _{1}, \dots ,\alpha _{k}, \beta _{1}, \dots ,\beta _{s})’$, and for the generalized logit model let $\bbeta =(\alpha _{1}, \ldots ,\alpha _{k}, \bbeta _{1}’,\ldots , \bbeta _{k}’)’$. The gradient vector and the Hessian matrix are given, respectively, by



where $l_ j=\log L_ j$is the log likelihood for the jth observation. With a starting value of $\bbeta ^{(0)}$, the maximum likelihood estimate ${\widehat{\bbeta }}$of $\bbeta $is obtained iteratively until convergence is obtained:

\[  \bbeta ^{(m+1)} = \bbeta ^{(m)} - \bH ^{-1}\mb{g}  \]

where $\bH $and $\mb{g}$are evaluated at $\bbeta ^{(m)}$. If the likelihood evaluated at $\bbeta ^{(m+1)}$is less than that evaluated at $\bbeta ^{(m)}$, then $\bbeta ^{(m+1)}$is recomputed by step-halving or ridging.

The covariance matrix of ${\widehat{\bbeta }}$is estimated by

\[  \widehat{\mbox{Cov}}({\widehat{\bbeta }})= \widehat{\bI }^{-1}  \]

where the observed information matrix $\widehat{\bI }=-\widehat{\bH }$is computed by evaluating $\bH $at ${\widehat{\bbeta }}$.

#### Firth’s Bias-Reducing Penalized Likelihood

Firth’s method is currently available only for binary logistic models. It replaces the usual score (gradient) equation

\[  g(\beta _ j) = \sum _{i=1}^ n (y_ i-\pi _ i)x_{ij} = 0 \quad (j=1,\ldots ,p)  \]

where p is the number of parameters in the model, with the modified score equation

\[  g(\beta _ j)^* = \sum _{i=1}^ n \{  y_ i-\pi _ i +h_ i(0.5-\pi _ i)\} x_{ij} = 0 \quad (j=1,\ldots ,p)  \]

where the $h_ i$s are the ith diagonal elements of the hat matrix $\bW ^{1/2}\bX (\bX ’\bW \bX )^{-1}\bX ’\bW ^{1/2}$and $\bW = \mbox{diag}\{ \pi _ i(1-\pi _ i)\} $. The Hessian matrix is not modified by this penalty, and the optimization method is performed in the usual manner.

### Existence of Maximum Likelihood Estimates

The likelihood equation for a logistic regression model does not always have a finite solution. Sometimes there is a nonunique maximum on the boundary of the parameter space, at infinity. The existence, finiteness, and uniqueness of maximum likelihood estimates for the logistic regression model depend on the patterns of data points in the observation space (Albert and Anderson, [1984](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_references.htm#statug_logisticalbe_a84); Santner and Duffy, [1986](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_references.htm#statug_logisticsant_t86)). Existence checks are not performed for conditional logistic regression.

Consider a binary response model. Let $Y_ j$be the response of the jth subject, and let $\mb{x}_ j$be the vector of explanatory variables (including the constant 1 associated with the intercept). There are three mutually exclusive and exhaustive types of data configurations: complete separation, quasi-complete separation, and overlap.

**Complete Separation**

There is a complete separation of data points if there exists a vector $\mb{b}$that correctly allocates all observations to their response groups; that is,

\begin{eqnarray*}  \left\{  \begin{array}{ll} \mb{b}’\mb{x}_ j > 0 &  Y_ j = 1 \\ \mb{b}’\mb{x}_ j < 0 &  Y_ j = 2 \end{array} \right. \end{eqnarray*}

This configuration gives nonunique infinite estimates. If the iterative process of maximizing the likelihood function is allowed to continue, the log likelihood diminishes to zero, and the dispersion matrix becomes unbounded.

**Quasi-complete Separation**

The data are not completely separable, but there is a vector $\mb{b}$such that

\begin{eqnarray*}  \left\{  \begin{array}{ll} \mb{b}’\mb{x}_ j \geq 0 &  Y_ j = 1 \\ \mb{b}’\mb{x}_ j \leq 0 &  Y_ j = 2 \end{array} \right. \end{eqnarray*}

and equality holds for at least one subject in each response group. This configuration also yields non­unique infinite estimates. If the iterative process of maximizing the likelihood function is allowed to continue, the dispersion matrix becomes unbounded and the log likelihood diminishes to a nonzero constant.

**Overlap**

If neither complete nor quasi-complete separation exists in the sample points, there is an overlap of sample points. In this configuration, the maximum likelihood estimates exist and are unique.

Complete separation and quasi-complete separation are problems typically encountered with small data sets. Although complete separation can occur with any type of data, quasi-complete separation is not likely with truly continuous explanatory variables.

The LOGISTIC procedure uses a simple empirical approach to recognize the data configurations that lead to infinite parameter estimates. The basis of this approach is that any convergence method of maximizing the log likelihood must yield a solution giving complete separation, if such a solution exists. In maximizing the log likelihood, there is no checking for complete or quasi-complete separation if convergence is attained in eight or fewer iterations. Subsequent to the eighth iteration, the probability of the observed response is computed for each observation. If the predicted response equals the observed response for every observation, there is a complete separation of data points and the iteration process is stopped. If the complete separation of data has not been determined and an observation is identified to have an extremely large probability ($\ge $0.95) of predicting the observed response, there are two possible situations. First, there is overlap in the data set, and the observation is an atypical observation of its own group. The iterative process, if allowed to continue, will stop when a maximum is reached. Second, there is quasi-complete separation in the data set, and the asymptotic dispersion matrix is unbounded. If any of the diagonal elements of the dispersion matrix for the standardized observations vectors (all explanatory variables standardized to zero mean and unit variance) exceeds 5000, quasi-complete separation is declared and the iterative process is stopped. If either complete separation or quasi-complete separation is detected, a warning message is displayed in the procedure output.

Checking for quasi-complete separation is less foolproof than checking for complete separation. The [NOCHECK](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticnocheck) option in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement turns off the process of checking for infinite parameter estimates. In cases of complete or quasi-complete separation, turning off the checking process typically results in the procedure failing to converge.

To address the separation issue, you can change your model, specify the [FIRTH](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticmfirth) option to use Firth’s penalized likelihood method, or for small data sets specify an [EXACT](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax15.htm) statement to perform an exact logistic regression.

### Model Fitting Information

For the jth observation, let ${\widehat{\pi }}_ j$be the estimated probability of the observed response. The three criteria displayed by the LOGISTIC procedure are calculated as follows:

* –2 log likelihood:

\[  -2\mbox{ Log L}=-2\sum _ j \frac{w_ j}{\sigma ^2} f_ j\log ({\widehat{\pi }}_ j)  \]

where $w_ j$and $f_ j$are the weight and frequency values of the jth observation, and $\sigma ^2$is the dispersion parameter, which equals 1 unless the [SCALE=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticscale) option is specified. For binary response models that use events/trials MODEL statement syntax, this is

\[  -2\mbox{ Log L}=-2\sum _ j \frac{w_ j}{\sigma ^2} f_ j [ \log {{n_ j}\choose {r_ j}} + r_ j \log ({\widehat{\pi }}_ j) + (n_ j-r_ j)\log (1-{\widehat{\pi }}_ j) ]  \]

where $r_ j$is the number of events, $n_ j$is the number of trials, ${\widehat{\pi }}_ j$is the estimated event probability, and the statistic is reported both with and without the constant term.

* Akaike’s information criterion:

\[  \mbox{AIC}=-2\mbox{ Log L}+2p  \]

where p is the number of parameters in the model. For cumulative response models, $p = k+s$, where k is the total number of response levels minus one and s is the number of explanatory effects. For the generalized logit model, $p = k(s+1)$.

* Schwarz (Bayesian information) criterion:

\[  \mbox{SC}=-2\mbox{ Log L}+p\log (\sum _ jf_ jn_ j)  \]

where p is the number of parameters in the model, $n_ j$is the number of trials when events/trials syntax is specified, and $n_ j=1$with single-trial syntax.

The AIC and SC statistics give two different ways of adjusting the –2 Log L statistic for the number of terms in the model and the number of observations used. These statistics can be used when comparing different models for the same data (for example, when you use the [SELECTION=](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm#statug.logistic.logisticselection) STEPWISE option in the [MODEL](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_syntax22.htm) statement). The models being compared do not have to be nested; lower values of the statistics indicate a more desirable model.

The difference in the –2 Log L statistics between the intercepts-only model and the specified model has a $p-k$degree-of-freedom chi-square distribution under the null hypothesis that all the explanatory effects in the model are zero, where p is the number of parameters in the specified model and k is the number of intercepts. The likelihood ratio test in the "Testing Global Null Hypothesis: BETA=0" table displays this difference and the associated p-value for this statistic. The score and Wald tests in that table test the same hypothesis and are asymptotically equivalent; for more information, see the sections [Residual Chi-Square](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details14.htm#statug.logistic.logisticreschitest) and [Testing Linear Hypotheses about the Regression Coefficients](http://127.0.0.1:60408/help/statug.hlp/statug_logistic_details36.htm).