

STAT0030统计计算  
评估3 (2018/19会议)

- 您的解决方案应该是您自己的工作，并通过电子方式提交  
Moodle将于2019年4月23日星期二下午4点之前发布。详细的提交说明如下如下。
- 延迟提交将被处以罚款，除非存在情有可原的情况（例如：  
医疗）由适当的文件支持。 罚款最新列出  
统计科学系学生手册的版本，可从  
部门网页。
- 未提交此课程评估将意味着您的整体考试  
标记记录为非完整，即您不会获得该课程的通行证。
- 任何抄袭或勾结通常会导致所有学生的分数为零 –  
volved，这也可能意味着您的整体考试成绩被记录为  
不完整的。 关于什么构成抄袭和勾结的指导原则可能是在  
部门学生手册中找到。 Turn-It-In抄袭检测  
系统可用于扫描您提交的剽窃或串通证据。
- 在统计考官会议确认之前，您的成绩将是临时成绩  
在2019年6月。
- 一般反馈将通过Moodle提供。

---

## 第2页

### STAT0030评估3 – 说明

- 1.您需要编写单个R函数。这个函数的代码应该保存在学生编号指定的.r文件中。例如，如果你的学生number为17101710，您的代码应保存在17101710.r文件中。
- 2.你的功能应该被彻底评论。它应该包含一个标题部分总结了逻辑结构，后面是函数的主体。主体本身应包含注释。
- 3.您需要提交以下内容：
  - R脚本的电子副本（见下文）。
  - 简要说明您的功能如何工作，以及它的摘要输出。解释应该包括，例如，任何数学的细节 – 在实施IWLS算法之前执行的计算。在哪里通过输出决定生产什么，你应该为这些决定辩护。作为粗略的指导，这个解释/总结长度不应超过2页（单面）。这也应该是以电子方式进行。
- 4.您的函数不应创建任何输出文件。
- 5.您的R功能和说明的电子副本应通过以下方式提交该课程的Moodle页面。查找标题为“使用此链接”的链接提交您的作业ICA3“并按照说明操作。

---

### 第3页

## STAT0030评估3 – R功能

假设 $Y$ 是几何随机变量的向量，其中 $Y_i \sim \text{Geo}(\pi_i)$

$$P(Y_i = y) = \pi_i (1 - \pi_i)^{y-1} \quad (y = 1, 2, 3, \dots),$$

为 $E(Y_i) = 1/\pi_i = 1/\text{率}\mu_i$ ，也就是说，和 $\text{VAR}(Y_i) = (1 - \pi_i)/\pi_i^2$  我。还假设 $x_i$ 是向量协变量，形成矩阵 $X$ 的第 $i$ 行，这样

$$\text{LN} \frac{1 - \pi_i}{\pi_i} = \ln[\mu_i - 1] = x_i^T \beta = \eta_i \text{我，比方说，}$$

对于某些系数向量 $\beta$ 。

这可以被视为GLM，因为几何分布是指数的家人和 $\eta_i$ 为 $\mu_i$ 的单调函数。

使用迭代加权最小二乘法写一个R函数来拟合这样的模型，并且检查拟合的型号。你的函数应该被称为`grm`（'几何回归模型'）。函数的参数应该是 $y$ ，一个使用`the`建模的响应向量如上所述的几何分布； $X$ ，协变量的设计矩阵和`startval`，模型系数的初始估计。如果用户没有提供`startval`值，

你应该提供一个默认值（例如一个零向量或任何其他明智的选择）或找到启动算法的其他方法。

您的函数应在没有用户干预的情况下运行，其值应为列表对象至少包含以下组件（如果您可以添加更多组件）觉得这些会有用）：

Y:	观察到的反应。
安装:	拟合值。
betahat:	估计的回归系数。
瑟伯塔:	估计回归系数的标准误差。
cov.beta:	估计回归系数的协方差矩阵。
电话号码:	在线性预测器中估计的系数的数量。
df.residual:	剩余的自由度。
偏差:	模型的偏差。

您的函数结构应类似于以下内容：

- 1.检查y和X的尺寸是否兼容，以及数据是否合适  
使用几何分布进行建模 – 如果没有，请使用适当的方法停止错误消息（欢迎您添加您认为可能需要的任何其他检查这里）。
- 2.执行IWLS过程以适合模型，并将结果输出到屏幕（如下面所描述的）。

---

## 第4页

- 3.生成残差图和其他适当的模型诊断。
- 4.将结果汇编到列表对象中，并将其作为函数的值返回。

在步骤2中，屏幕输出应包括：显示估计系数的表格客户，他们的标准错误，z统计和相关的p值；系数的数量估计；拟合模型的剩余自由度；以及对...的偏差拟合模型。如果您愿意，您可以输出任何其他相关信息。

在第3步中，您应该使用您对GLM的模型检查的知识来生成适当选择诊断。您不必生成与R相同的图当你绘制一个glm对象时。

你的函数不能使用glm命令（也不能使用像glm.fit这样的类似命令）！

## STAT0030评估3 – 提示

这个问题没有单一的“正确答案”。获得你需要的好标记  
明智地处理问题，并为你提供明确的理由  
这样做。对于清晰易读的代码，我们将给予信任。特别是代码  
评论不充分将受到处罚。

你应该确保你的功能产生明确和适当的输出。  
标记和格式化。

- 3.您不需要在此处分析任何数据; 但是，在标记此评估时 –  
您的功能将在一个或多个数据集上进行测试，以确保其正常运行  
正确。因此，您可能希望在一个简单的数据集上测试您的函数 –  
提交前，并可选择将您的测试脚本与您的功能一起提交  
如下面所描述的。
- 4.如果需要，您可以使用Workshop 8中的IWLS功能作为起点  
这个评估。
- 5.为了解释你的功能如何工作，你可能需要使用相当多的功能  
数学符号。我们鼓励您使用 $\text{L A T E X}$ 。话虽这么说，但是清晰可辨  
手写的解释也完全可以接受。
- 6.为了解释你的功能是如何工作的，你必须解释给定的  
分布在指数族中。
- 7.您的脚本将通过从假定的程序调用您的函数来进行测试  
你已经完成了问题的要求。这意味着，例如，你  
必须按照上面给出的顺序指定函数的参数，以及名称  
列表结果的各个元素必须与上面给出的相同。如果  
你不这样做，你的功能在被叫时会失败，你会失去分数。

---

## 第5页

- 8.R有一些与几何分布有关的内置程序。你可以用  
如果你认为它们会有用的话; 但是，请注意该定义  
R中的分布与上面给出的略有不同。
- 9.如果您还没有这样做，请阅读有关第一次ICA的一般反馈  
Moodle的。当ICA 2可用时，还要阅读有关ICA 2的反馈。

- 10.如果您遇到困难或需要建议，应该对此评估提出疑问  
在办公时间或在moodle论坛上制作。有关办公室的详细信息  
小时，以及预约的链接，请参阅Moodle页面。
- 11.您可能在某些时候发现使用约定 $0 \log 0 = 0$ 很有用。严格来说  
说这个数量是未定义的，但由于 $\varepsilon \log \varepsilon \rightarrow 0$ 为 $\varepsilon \rightarrow 0$ ，所以它是a  
在许多情况下普遍接受的规则，并且可以在这个练习中很好地为您服务。

### STAT0030评估3 – 可选的测试用例脚本

您可以编写第二个脚本来加载数据集，适合回归模型  
使用您的grm实现，并输出一系列估计和诊断。  
数据的选择属于您自己，但执行必须由您的任何用户重现  
脚本。因此，将自己局限于可以从R包加载的数据集，或者哪些数据集  
可以从脚本本身的R代码构造。对于前者，我们建议使用  
包数据集。您可以选择数据和输出。这个脚本的目标  
是为了向我们展示你的脚本在实践中工作的一个例子，万一我们  
在我们自己的测试用例上运行它有任何问题。例如，如果您的脚本有效  
正确地使用您提供的数据，但不是我们的所有测试用例，我们将能够  
为演示脚本的工作情况给予适当的信任。对于  
但是，我们要求您的测试用例脚本写得清楚  
评论说。只要代码清晰且可重复，格式就由您自己决定。

如果您使用此选项，请将测试脚本作为第二个文件上载。如果你的学生  
数字是17101710，比方说，请使用格式17101710test.r。

### STAT0030评估3 – 标记指南

该评估标记为50个。标记大致细分为以下内容  
组件：11个标记用于正确实现IWLS算法，21个标记用于  
正确检查输入，正确显示输出，以及良好的编码风格，  
和18个标记清楚说明您的功能如何工作，以便正确诊断  
方差函数，偏差等的正确数学表达式。