

# Java 课程作业（选做）

## 1 $p$ -value

### 1.1 应用背景

在数据科学中，我们常常需要利用观察到的数据来验证我们所作出的假设是否是合理的，这在统计学中称为“假设检验（Hypothesis Testing）”。那么，如何验证假设的合理性呢？熟悉统计的人会想到“显著水平（**significance level**）”和“ **$p$ -value**”等概念。表 1 给出了一个利用  $p$ -value 进行假设检验的例子。

**表 1  $p$ -value 的重要性：一个利用  $p$ -value 进行假设检验的例子**

——自然界的所有差异，换来了整个自然界的平静（英·蒲柏）。

在下面的文献（p.3, sec. 2）

Tzimiropoulos, G., S. Zafeiriou and M. Pantic, *Subspace Learning from Image Gradient Orientations*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2012. 34(12): p. 2454-2466. ([.pdf](#))



中，作者提到了：

“自然界中的任意两幅不同的图像的梯度方向之差服从均匀分布”

这一假设（null hypothesis）。那么，怎么说明这一假设的合理性呢？作者写道：

The null hypothesis was tested using the **Kolmogorov-Smirnov test**. For a **significance level** equal to 0.01, the null hypothesis was accepted for 94.05% of the image pairs with mean  **$p$ -value** equal to 0.2848.

### 1.2 主要问题

1. 请认真阅读本节的全部参考资料；
2. 回答下列问题：
  - (1) 什么是  $p$ -value？
  - (2) 如何计算  $p$ -value？
  - (3) 与  $p$ -value 相关的主要概念，如：significance level、test statistic 等。
  - (4) 与  $p$ -value 相关的主要统计检验方法，如：chi-squared tests、Student's t-test、Kolmogorov-Smirnov test 等。
3. 编写 Java 程序实现  $p$ -value 的计算过程。

### 1.3 参考资料

- [1] <http://en.wikipedia.org/wiki/P-value>
- [2] Wasserman, L.A., All of statistics: a concise course in statistical inference. 2004: Springer. ([.pdf](#), ch10, p.165~p.189)

### 1.4 考核方式

1. 以 PPT 的形式进行口头报告，能够简单、清晰地描述  $p$ -value（18 分）；

2. Java 代码实现正确（7 分）。
- 

## 2 Gibbs 采样

### 2.1 主要问题

- (1) 什么是 Gibbs 采样？认真阅读参考资料[1]和[2]，弄懂 Gibbs 采样以及与 Gibbs 采样相关的主要问题，如：其中的 MCMC 算法（Markov chain Monte Carlo (MCMC) algorithm）等。
- (2) 用 Java 实现 Gibbs 采样及其所涉及的主要算法。

### 2.2 参考资料

- [1] [http://en.wikipedia.org/wiki/Gibbs\\_sampling](http://en.wikipedia.org/wiki/Gibbs_sampling)
- [2] <http://blog.csdn.net/yangliuy/article/details/8457329>

### 2.3 考核方式

1. 以 PPT 的形式进行口头报告，能够简单、清晰地描述 Gibbs 采样（18 分）；
  2. Java 代码实现正确（7 分）。
- 

## 3 RollCall

### 3.1 主要问题

关于 RollCall 的详细资料见我的工大主页：  
[http://www.tdc.zjut.edu.cn/UTADB/teacher/search\\_one\\_intro.jsp?teacher\\_id=04922](http://www.tdc.zjut.edu.cn/UTADB/teacher/search_one_intro.jsp?teacher_id=04922)，在【应用软件】部分提供了：[RollCall 的使用说明.pdf](#) 及其 [C++源码](#)。目前，RollCall 尚不完善，尤其是如下三个功能模块，亟待实现，请用 C++ 或 Java 实现之（注：只要实现任一个模块，就可以拿到 25 分，抵消考勤成绩）：

1. 记录功能。（25 分）目前只有“点名”的功能，缺乏“记录”的功能，我们将在随后的版本中添加：
  - 1) 考勤记录：到、未到两种状态，默认为“到”。
  - 2) 点名记录：所有被点名者被点名的时间、次数、表现情况。
2. 辅助功能。（25 分）
  - 1) 教学日历：可以很方便地导入、编辑、生成、查看教学日历，并且可以很方便地从教学日历中查看考勤记录和点名记录。
  - 2) 声音提示：被点到者的名字可以用语音提示。
3. 用户界面。（25 分）

- 1) 数据刷新：当新增加班级数据时，不需要重启软件，可以直接在界面中点击刷新。
- 2) 随机抽样的频率定制：用户可以很方便地调整在一次随机点名中，随机抽样的频率。
- 3) 头像显示：每次随机抽取时，要求在屏幕上滚动的不仅是学号或名字，还有头像。

### **3.2 考核方式**

1. 以软件能否正确运行为唯一的考核方式。