2015中德合作上海小学生运动检测项目设计说明

# 概述

本项目长期目标为构建上海全年级小学生运动素质检测平台，包括完备的数据采集手段、可靠并持续优化的评价体系。

当前目标包括：

* 可录入各运动项目评价参数
* 可登记学生基本信息（逐条/批量）
* 可登记原始测试数据（逐条/批量）
* 根据评价参数，计算并输出结果

主要操作界面为Web，批量导入文件格式为Excel或CSV。

# 分析

评价体系是本项目需要重点维护的部分，具体来说就是将学生的运动测试（原始）数据转化为标准评分。

根据现有资料，采用正态分布模型作为核心计算模型，计算某个数值的累计概率（即小于等于该数值的可能性），作为运动能力参考值（标准评分）。

原始的单项运动测试数据通常多个数据项，需要经过处理（合计、平均等）为单个数值，从而可以在后续概率计算中使用。

正态分布模型的准确性是至关重要的，通常认为**大量、随机**样本是其基础，所以有以下几点值得关注：

* 评价模型应与评价的对象相匹配
  + 不能用德国的模型评价中国的学生
  + 不能用上海的模型评价全中国的学生（反之是可以的）
* 评价模型的修正依赖于合理的样本
  + 若需要获得全国的模型，那么用于推算这个模型的样本数据不能局限为上海。
  + 对于样本的随机性可能需要人工验证。
  + 系统自动根据现有数据修正模型时，首先确保样本规模不能小于某个下限。

对于每一项测试评价模型，不同的性别、月龄均构成了一个独立分组。其中性别差异很容易理解，不再说明。月龄是测试对象当时的年龄（按月计算），原因是儿童身体处于快速发展时期，即使相差数月相应的测试数据也会发生很大变化。根据现有的评价模型（按照正态曲线方式呈现），会发现随着年龄变化，正态曲线也是连续变化的。我们甚至可以认为存在一种包含年龄因素的分布函数（评价某个具体测试数据时，只需要固定年龄因素即得到正态分布模型）。

# 模型

以下定义的模型根据当前目标设计，省略了一些非必要数据项。

## 分布因素(Factor)

**说明：**

运动检测项目的评价参数，不同的性别、月龄、测试点均构成一个分组。

**属性：**

|  |  |  |
| --- | --- | --- |
| 名称 | 标识 | 说明 |
| 性别 | gender | 男、女 |
| 月龄 | month\_age | 整数（间隔为1） |
| 测试点 | movement\_type | 20米跑、后退平衡、侧向跳… |
| 平均值 | mean | 正态分布参数 |
| 标准偏差 | standard\_deviation | 正态分布参数 |

## 学生信息(Student)

**说明：**

学生的基本信息

**属性：**

|  |  |  |
| --- | --- | --- |
| 名称 | 标志 | 说明 |
| 姓 | last\_name | 中文 |
| 名 | first\_name | 中文 |
| 性别 | gender | 男、女 |
| 出生日期 | birth\_date |  |
| 学校名称 | school\_name | 文本  后续版本可能调整为学校Ref |
| 班级名称 | class\_name | 文本  后续版本可能调整 |

## 测试原始记录(TestRefData)

**说明：**

测试现场采集的数据

**属性：**

|  |  |  |
| --- | --- | --- |
| 名称 | 标志 | 说明 |
| 测试日期 | testing\_date |  |
| 测试编号 | testing\_number | 整数 |
| 测试学生 | student | 学生Ref |
| 身高 | height | 测试当天的身高 |
| 体重 | weight | 测试当天的体重 |

## 测试原始记录单项数据(TestRefDataItem)

**说明：**

从属于测试原始记录

**属性：**

|  |  |  |
| --- | --- | --- |
| 名称 | 标志 | 说明 |
| 测试原始记录 | test\_ref\_data | 测试原始记录Ref |
| 测试点 | movement\_type | 20米跑、后退平衡、侧向跳… |
| 数据项 | key | 1、2、3，可能值因测试点不同而改变 |
| 数据值 | value | 测试结果 |

## 测试总结(TestSummaryData)

**说明：**

经过处理的测试记录

**属性：**

|  |  |  |
| --- | --- | --- |
| 名称 | 标志 | 说明 |
| 测试学生 | student | 学生Ref |
| 测试日期 | testing\_date |  |
| 身高 | height | 测试当天的身高 |
| 体重 | weight | 测试当天的体重 |
| 月龄 | month\_age | 根据测试日期和出生日期推算 |
| 日龄 | day\_age | 根据测试日期和出生日期推算 |
| 测试原始记录 | test\_ref\_data | 测试原始记录Ref |

## 测试总结单项数据(TestSummaryDataItem)

**说明：**

从属于测试总结

**属性：**

|  |  |  |
| --- | --- | --- |
| 名称 | 标志 | 说明 |
| 测试总结 | test\_summary\_data | 测试总结Ref |
| 测试点 | movement\_type | 20米跑、后退平衡、侧向跳… |
| 数据值 | value | 处理后的测试数据（根据原始测试数据） |
| 评估时间 | evaluate\_date | 本次评估执行时间 |
| 评估数值 | evaluate\_value | 评估分值  注意，若后续调整了评估模型（分布因素），那么此数值可能处于过期状态。 |

# 技术框架

本项目为小型研究型项目，主要工作方向为科学计算。经过比较，选择Python为主要开发语言。

*Python是一种面向对象的、动态的程序设计语言，具有非常简洁而清晰的语法，既可以用于快速开发程序脚本，也可以用于开发大规模的软件，特别适合于完成各种高层任务。*

*随着NumPy、SciPy、matplotlib、ETS等众多程序库的开发，Python越来越适合于做科学计算。与科学计算领域最流行的商业软件MATLAB相比，Python是一门真正的通用程序设计语言，比MATLAB所采用的脚本语言的应用范围更广泛，有更多程序库的支持，适用于Windows和Linux等多种平台，完全免费并且开放源码。虽然MATLAB中的某些高级功能目前还无法替代，但是对于基础性、前瞻性的科研工作和应用系统的开发，完全可以用Python来完成。*

同时选择Django，这是一个轻量级Web框架，较为完备，并且可以部署在几乎任何服务器环境上。

数据库采用其默认的sqlite，便携型是其主要优势。另外，因为Django采用了根据模型自动维护数据的方式，所以后续若有性能压力，可以很容易迁移到其它数据库。

# 操作模式

1. 录入数据（对于单个学生数据，记录首次录入时间）
2. 计算数据（计算时检查数据有效性，报告数据此时确定）
3. 查看数据（不进行任何数据更新操作）

# 对照表

## 测试点(MovementTypes)

**说明：**

代码的选择参照了导出的Excel数据表，某些看起来不容易辨识。

|  |  |  |  |
| --- | --- | --- | --- |
| 英文 | 中文 | 代码 | 备注 |
| 20m-sprint | 20米冲刺跑 | 20m |  |
| Balancing | 平衡 | bal |  |
| Sidesway Jumping | 侧向跳 | shh |  |
| Trunk-bend | 直身前驱 | rb |  |
| Ball throwing | 投掷 | ball |  |
| Push-ups | 俯卧撑 | ls |  |
| Situps | 仰卧起坐 | su |  |
| Stand Long Jump | 跳远 | sws |  |
| 6-min Run | 六分跑 | lauf |  |