

对足球运动员身体素质及成长性评价体系的研究

摘 要

足球运动员身体素质的好坏决定其个人的发展及球队的整体水平，而体能测试结果直接体现了球员的身体素质。本文在对足球运动员各位置的特点和相应数据做出详细分析的基础上，针对现有体能测试存在的问题，对不同位置的运动员给出了与其相适应的合理权重，得出了更合理的体测标准，建立了体能成长性评估模型，实现了基于球员当前测试数据及年龄的成长性评估。

针对问题一，我们首先对数据进行了预处理，按旧标准得出每位运动员体测成绩，按给定合格线给出不合格球员名单及其成绩，并对每队球员成绩求均值得出各球队综合体能测试成绩。

针对问题二，我们通过问题一中所求的成绩做出了总体成绩分布的频率直方图，并借此求得反映总体球员前 85% 水平的成绩，定为新的合格线，并给出新合格线下无法选满 22 人的球队名单。

针对问题三，经查阅文献初步得出，对球场上不同位置的球员，各方面身体素质的要求应该是不同的；通过对守门员成绩的单独分析，即按问题二中合格线算出合格率，发现其低于总体水平。结合以上两点可判断原有权重标准不够合理。对此我们建立了组合权重计算模型，以获得各测试项目的更合理的权重。首先我们使用层次分析法，收集专家填写的问卷中的数据，构建对于不同位置的各体能因子的重要性比较矩阵，进而计算出其主观权重，进而得出体能因子对应项目的权重；其次，我们采用熵权法计算不同位置球员在各个指标下的成绩的信息熵，进而计算出客观权重，最后我们采用有序加权平均算子，对主观权重和客观权重进行加权得到最终的组合权重。

针对问题四，首先，我们通过测试收集到的数据，基于统计学的方法，画出了某一特定的测试项目的成绩在运动员群体中的数量分布散点图，发现其近似服从高斯分布，于是我们用高斯分布概率密度函数进行拟合，再计算出等间距的 100 个分位数，从而制定出了新的测试标准；其次，我们把身高、体重、体脂等身体形态指标也加入了运动员的身体素质评估体系中，丰富了第三问中所建立的综合身体素质层次框架，并计算出了新的权重；最后，我们收集了一份带有年龄信息的职业足球运动员全方位信息数据集，提取出其中的身体素质数据，利用多项式拟合了这些身体素质指数与年龄之间的函数关系，建立了球员成长性评估模型。基于一个球员当前的年龄和身体素质数据，我们可以通过成长性评估模型，计算出他未来所能达到的巅峰身体素质水平，从而评估他的成长空间。

该评价模型对足球运动员身体素质及成长性评价体系的改进有重要意义。

关键词：层次分析法 熵权法 有序加权平均算子 多目标评价 高斯分布拟合 多项式拟合

一、问题重述

足球运动以技术为核心，体能为基础。运动员身体素质的好坏决定其个人的发展及球队的整体水平。体能测试是检测和选拔运动员的一项重要测试。

不同位置的足球运动员身体素质有所差异，各个年龄段的球员体能素质也有不同，应从发展的眼光去看待。如何优化足球体能测试，建立新标准，从而使测试选拔更合理，将是本文探讨的重点。

问题一，对 15 个球队体能测试的数据按原有标准权重进行处理，得出个人和球队成绩，按原有合格线（6 分，满分 10 分）选出合格球员。

问题二，按所有参加测试球员成绩的 85%，确定新的合理的合格线，列出新合格线下无法选满 22 人的球队。

问题三，说明不同位置球员的体能测试按原有同一权重标准评分是否合理；试根据足球不同位置的特征对每个位置建立新的权重标准，说明其合理性。

问题四，说明目前测试方法否从成长发展的角度反应球员体能水平，如何改进测试及参数使其满足要求，做出模型并解释合理性。

二、问题分析

2.1 问题一的分析

问题一主要任务可分为两大部分：数据的处理和数据计算

针对数据处理，题目所给数据集中数据有数据大量缺失、数据少量缺失、数据与实际不符、数据重复等问题：

（1）针对大量缺失数据，使用其他数据进行替代将影响到公平性，故在计算个人综合成绩时将其记为 0，在计算组内平均成绩时将其删去；

（2）对于数据少量缺失，因为缺失数据在总成绩中占比较小，且去除该项成绩后该球员依旧能够通过测试，故用参加测试所有该位置球员该项测试的平均成绩替代此成绩；

（3）对于数据与实际不符的情况，推测其产生的原因并对进行合理的填补；

（4）针对数据重复的问题，由于不确定其产生原因，最终决定不对其进行处理。

之后可通过计算求解出问题一。

2.2 问题二的分析

问题一中计算的所有个人成绩分布在 $[0, 10]$ 这一区间内，其中最大值为 9.85，最小值为 0，我们取组距为 0.5，将 $[0, 10]$ 划分为 20 个小区间（20 个组），计算各组相应的频率 f 并将其除以组距作为纵坐标，做出直方图（间），图形总面积为 1。假设合格线取 a 时，直方图上高于 a 的区间的部分的面积为 0.85，计算出 a ，即该成绩可在一定程度上反应该样本总体 85% 的水平，即为所求合格线。再将 1 中成绩与其比较，即可得出无法选满 22 人的球队。

2.3 问题三的分析

按第二问求出的合格线，每个球队均有通过测试的守门员，但对所有参加测试的 43 名守门员成绩统计后，有 31 人通过，占比 72.09%，低于总体通过率 85%，可以认为该体测标准对守门员来说较严格。

对于不同位置的球员，比赛时对各种不同方面的身体素质要求不同。比如对于前锋，在比赛过程中需要突破对手的防守带球射门，在进攻过程中常常需要灵活地躲避，快速地运球，这就对前锋球员的速度与灵敏性提出了更高的要求；而对于中场球员，在比赛中需要在进攻和防守转换的情况下，在赛场上来回跑动，这对中场球员的耐力就提出了更高的要求。

由此可见，对球场上不同位置的球员，各方面身体素质的要求应该是不同的，而各方面的身体素质水平可以体现在各种项目的测试成绩中，因而对于一个球员的身体素质测试，应该对于不同的测试项目赋予不同的权重，故问题一中的权重是不合理的。

我们通过建立数学模型来确定这些权重。首先，对于不同位置球员不同项目成绩的相对重要性往往是定性的、模糊的，而我们需要的是一个定量的、精确的权重，因此，拟采用层次分析法（AHP）对定性问题进行定量的求解；其次，评分的目的是为了区分，我们收集到了一些球员的身体素质数据，因此我们可以采用熵权法（EWM），对于某一指标，球员们成绩差异越大，代表着该指标包含的信息越多，区分度也就越大，从而可以设置较高的权重；最后，我们希望将主观权重与客观权重组合起来，以克服两种赋权方法各自的缺点，获得更合理的权重。

当然，这样的组合权重并不一定真的“合理”，在得到结果后，我们会对结果做进一步的分析，并优化我们的模型。

2.4 问题四的分析

首先，根据运动员身体素质的成长规律，在不同的年龄时，运动员的体能水平并不相同。因此从球员的成长的角度出发，我们不仅应该关注球员在测试当下的体能评分，还应该关注他在未来的时候身体素质所能达到的最好状况，以及什么时候能够达到这一巅峰水平。

其次，通过文献阅读与专家访谈，我们发现，在对运动员的体能评估中，身高、体重、体脂等身体形态指标也是非常重要的一部分。而且对于足球场上的不同位置，对足球运动员的身体形态要求也不一样。比如对于守门员，其身高较高一些会更利于球门处的防守；而对于后卫，在比赛过程中会有经常性的身体对抗，而在体重保持在合理范围的前提下，体重越高，越有利于对抗，运动员的身体素质也应该越好。而当前的测试方法中并无对于身高、体重等指标的测试，因此目前的测试方法有一定的改进空间。

最后，本题中使用的测试标准参考了国家体科所 2020 年相关研究（但仅有 U18 年龄组标准），和个别俱乐部测试结果（U20 年龄组）。但是，有些测试项目找不到大量的数据支撑或过于简单（如箭头跑、T 形跑），标准不见得“标准”。而通过测

试收集到的数据，基于统计学的方法，我们可以估计出对于某一特定的测试项目，该项目的成绩在运动员群体中的数量分布，从而可以制定出相应的标准。收集到的数据越多，基于统计与回归进行的拟合也就越准确，最终制定出的标准就越合理。

三、模型假设

- 1) 专家具有足够的权威性，且对问卷中问题的回答都是基于其专业知识储备。
- 2) 某一运动员在某一年龄体能素质所占其巅峰时期比例与体能素质平均曲线所反映比例相同。
- 3) 某一特定的测试项目的成绩在运动员群体中的数量分布可用高斯分布拟合。
- 4) 身体素质指数与年龄之间的函数关系可用多项式函数拟合。

四、符号说明

符号	说明	单位
CI	一致性指标	
RI	平均随机一致性指标	
CR	一致性比例	
A^k	比较矩阵， $k = 1, 2, 3, 4$ 分别代表前锋、中场、后卫、守门员	
a_{ij}	第 i 个因子与第 j 个因子的重要性之比	
W^S	主观权重， W 代表权重， s 代表主观	
W^O	客观权重， W 代表权重， o 代表客观	
e_j	第 j 个指标的信息熵， $j=1, 2, \dots, m$	

五、模型的建立与求解

5.1 问题一模型的建立与求解

5.1.1 数据处理

首先需要对问题一中所给数据进行处理：

对于缺失少量数据缺失的队员：

表 5.1 缺失少量数据队员

球员编号	30m 跑	箭头跑	立定跳远	纵跳	引体向上	YOYO	位置	球队
7	3.94		2.45	49.4	19	1080	中场	11
31	3.93	5.31	2.76	59.2	10	1040		11
8	3.93	7.95	2.76	61.6	11	8840	中场	13

对于 11 组中编号 7 的球员，其只缺少箭头跑数据，按所给权重占总成绩 15% 算，如果取 6 分为及格线，该队员即使将该缺失数据按最低 0 分来算，其总分仍有

7.05 > 6 为合格球员，我们决定将其缺失数据按其同位置所有测试球员的均值代替，这样对其个人能否入选合格球员及对球队综合成绩的影响最小；13 组 8 号只有 YOYO 一项数据明显偏离，与其他球员数值对比，其记录错误概率最大，实际计算按 840 处理。而编号 31 只缺少位置对问题一暂无影响。

对于缺失大量数据或大量数据不一致球员：

表 5.2 缺失大量数据或大量数据不一致的球员

球员编号	30m 跑	箭头跑	立定跳远	纵跳	引体向上	YOYO	位置	球队
7				0			前锋	6
3				0			中场	7
7				0			前锋	10
18	4.18	8.34	0	0	0	0	后卫	12
22							后卫	15

如上图列出了样本中的所有情况，对于缺失量太大，或不一致量太多，如果采用上述均值代替的方法等于这些球员 100% 入选合格队员，风险较大并对其他球员不公平；对于图中 18 号，虽然有两项数据，但由于各项数据所测能力不同，其间几乎没有关联，采用拟合各项数据关系来填补所缺数据显然是不合理的，而采用均值代替也等于该球员直接入选合格，风险较大也缺乏公平性。综上所述，对于这些球员数据我们直接将其从样本中删除（总体成绩计算时不考虑，个人成绩记为 0），将其列入不合格球员中，这样保证了所选队员都是符合要求的，对于其他球员是公平的；对于其个人，因设备或其他因素造成数据缺失或不一致，我们没有更合理的方法对其做出处理。

此外，第六支队伍与第十支队伍的数据完全相同，明显与正常情况不符，但无法确定是哪一支队伍成绩出现了错误，因此最终决定不对该成绩进行处理。

5.1.2 数据计算

a) 通过 python 编写程序将每个队员项目测试数据按所给标准换算为每项成绩，并乘上相应权重得出每个运动员的最终体能综合成绩，具体见附录

b) 通过 python 编写程序在 a 问的基础上筛选出每个球队不合格球员名单及成绩，具体数据见下表：

表 5.3 每个球队不合格球员名单及成绩

队伍	队员 编号	综合 得分	队伍	队员 编号	综合 得分	队伍	队员 编号	综合 得分
1	9	5.7	6	9	3.55	10	8	5.4
1	13	4.45	6	10	5.8	10	9	3.55
1	22	5.45	6	11	4.7	10	10	5.8
2	4	5.55	6	13	5	10	11	4.7
3	21	4.5	6	14	5.8	10	13	5
4	3	5.6	6	15	3.85	10	14	5.8
4	4	4.9	6	16	4.85	10	15	3.85
4	16	5.9	6	18	3.45	10	16	4.85
4	20	5	6	19	5.3	10	18	3.45
4	21	5.4	6	20	4.7	10	19	5.3
4	22	5.85	6	21	2.1	10	20	4.7
5	1	5.25	6	22	5.45	10	21	2.1
5	2	5.45	6	23	3.4	10	22	5.45
5	7	4.7	6	24	4.8	10	23	3.4
5	11	5.85	6	25	5.7	10	24	4.8
5	13	5.9	7	3	0	10	25	5.7
5	14	5.1	7	12	5.9	12	2	5.6
5	23	5.4	7	18	3.7	12	5	4.5
5	27	4.6	7	21	5	12	11	5.45
5	28	5.15	9	1	5.05	12	17	5.55
6	1	3.05	9	19	5.05	12	18	1.8
6	2	4.2	10	1	3.05	12	23	5.55
6	3	5.9	10	2	4.2	15	11	5.3
6	4	3.85	10	3	5.9	15	12	5.8
6	5	4.9	10	4	3.85	15	22	0
6	7	0	10	5	4.9	15	30	4.25
6	8	5.4	10	7	0			

c) 综合得分=每个球队有效成绩总和/球队有效人数，具体结果见下表：

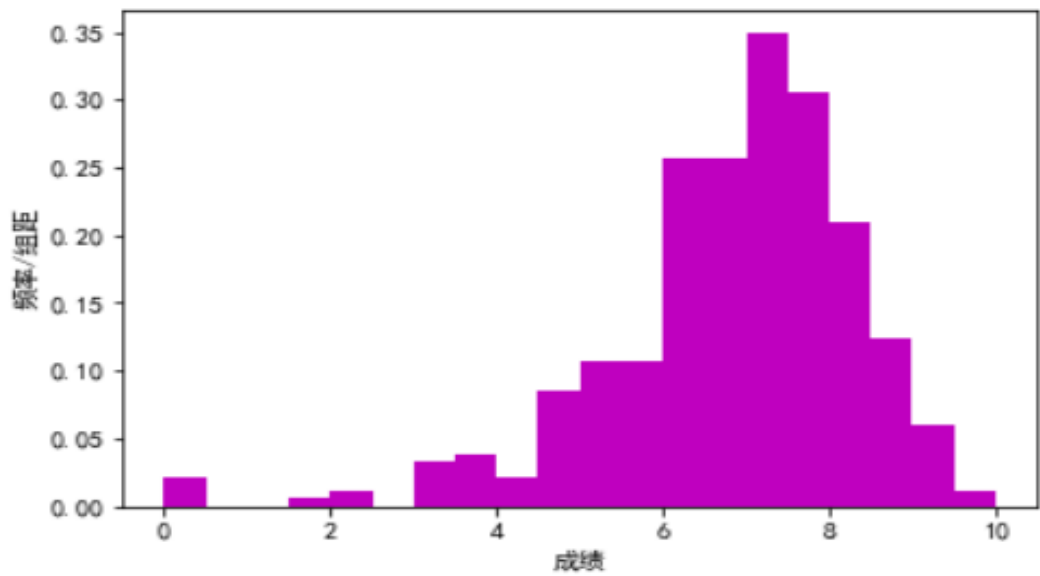
表 5.4 队伍综合得分

队伍编号	队伍综合得分	队伍编号	队伍综合得分
1	6.6083	9	6.9477
2	7.2442	10	4.7583
3	7.0045	11	7.8065
4	6.7609	12	7.0068
5	6.3286	13	8.3917
6	4.7583	14	6.75
7	6.945	15	7.6621
8	8.0444		

5.2 问题二模型的建立与求解

用 python 统计各区间人数换算为频率，用 matlab 画出直方图如下所示：

图 5.1 各区间人数直方图



由直方图得：

$$0.85=(20+48+48+65+57+39+23+11+2)/373+(5.5-a)*(20/373)/0.5 \tag{1}$$

可得 $a=5.39875$

即所求合理合格线为 5.39875

按此为合格线，则无法挑选出 22 人的球队有：3 号球队、4 号球队、6 号球队、7 号球队、9 号球队、10 号球队、12 号球队和 14 号球队。具体如下表所示：

表 5.5 各球队合格人数

队伍编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
合格人数	23	26	21	21	23	9	18	27	20	9	31	21	24	21	28

5.3 问题三模型的建立与求解

5.3.1 主观权重计算模型的建立与求解

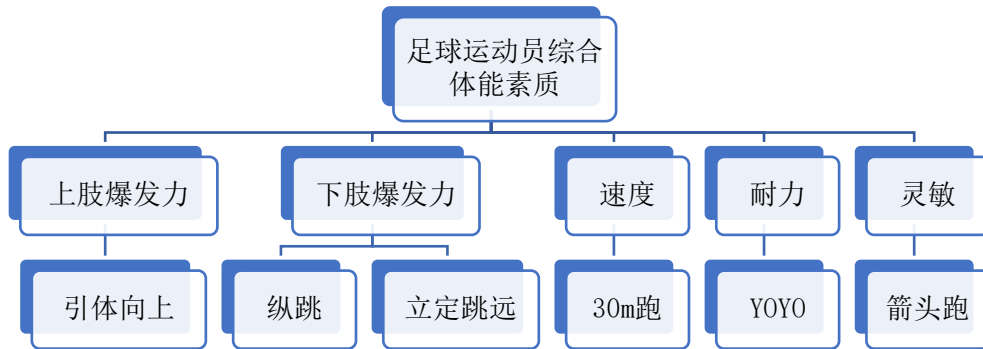
采用层次分析法（AHP）计算各个项目的主观权重。

步骤一：建立指标层次结构

无论什么运动的体能素质，都离不开速度、力量、协调、柔韧、灵敏、耐力等方面，足球也不例外，各项测试之所以能够作为评定球员身体素质的指标，也是因为它们在一定程度上反映了球员的速度、力量等情况。通过查阅相关文献并结合题目所给测试特点，本文选取了上肢爆发力、下肢爆发力、耐力、速度和灵敏五个方面作为球员身体素质的评价标准，并与六项测试相互对应，其关系如下：

通过分别研究这五项身体素质对各个位置球员的重要性，即可求得各项测试对各个位置球员的权重，因此建立指标层次结构如下：

图 5.2 体能素质指标层次结构



步骤二：构造成对比较阵

对于不同位置的球员，各种体能评估指数对运动员的重要程度不同。所以对于不同位置的球员，成对比较阵不同。本文通过让专家对各种项目的重要程度进行相互比较的方式，以问卷收集的方式获取成对比较矩阵。记 $A^k(k=1,2,3,4)$ 分别为前锋、中场、后卫、守门员各项目重要性的成对比较阵， $A^k = (a_{ij})_{l \times l}$ ，其中 m 为体能素质因子个数， a_{ij} 为第 i 个因子与第 j 个因子的重要性之比。

$$a_{ij} = \frac{1}{a_{ji}}, a_{ij} > 0, i, j = 1, 2, \dots, l, (a_{ii} = 1) \quad (1)$$

经过专家填写问卷，我们得到的比较矩阵如下：

表 5.7 体能素质比较矩阵

体能 素质 因子	守门员					后卫				
	上肢	下肢	速度	耐力	灵敏	上肢	下肢	速度	耐力	灵敏
上肢	1	2/3	5/6	5/6	4/5	1	2/3	1/2	1/3	3/4
下肢	3/2	1	5/4	5/4	6/5	3/2	1	1	2/3	6/5
速度	6/5	4/5	1	1	5/6	2	1	1	2/3	6/5
耐力	6/5	4/5	1	1	5/6	3	3/2	3/2	1	4/3
灵敏	5/4	5/6	6/5	6/5	1	4/3	5/6	5/6	3/4	1

体能 素质 因子	中场					前锋				
	上肢	下肢	速度	耐力	灵敏	上肢	下肢	速度	耐力	灵敏
上肢	1	1/3	1/5	1/5	1/3	1	1/5	1/5	1/5	1/3
下肢	3	1	5/6	2/3	1	5	1	1	5/4	5/4
速度	5	6/5	1	5/6	5/4	5	1	1	5/4	5/4
耐力	5	3/2	6/5	1	3/2	3	4/5	4/5	1	1
灵敏	3	1	4/5	2/3	1	3	4/5	4/5	1	1

步骤三：不一致性检验

计算一致性指标 CI (consistency index)用来衡量 A 的不一致程度：

$$CI = \frac{\lambda_{\max} - l}{l - 1}$$
 (2)

查找相应的平均随机一致性指标 RI (random index)：

表 5.8 随机一致性指标

1	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49

计算一致性比例 CR (consistency ratio)：

$$CR = \frac{CI}{RI}$$
 (3)

当 $CR < 0.1$ 时，认为矩阵 A 的不一致性是可以接受的，当 $CR \geq 0.1$ 时，则应修改比较阵 A 直至不一致性可接受为止。

一致性指标及一致性比例计算如图所示，经验证得，先前矩阵的不一致性均在可接受范围内。

表 5.9 体能素质一致性检验数据

位置	CI	CR
守门员	0	0
后卫	0.068	0.06
前锋	0.0033	0.003
中场	-0.0168	-0.015

步骤四：计算权重向量

求矩阵 $A = (a_{ij})_{l \times l}$ 的最大特征值所对应的特征向量并进行归一化，即可作为权重向量。在体能素质因子权重计算完成后，在每个因子下，重施上述步骤，计算各个项目的权重，然后同所属因子的权重分别相乘，就得到了各个项目的主观权重，其中 m 为测试项目的个数：

$$W^S = (w_1^S, w_2^S, \dots, w_m^S) \tag{4}$$

表 5.10 主观权重求解结果

位置	30m 跑	箭头跑	立定跳远	纵跳	引体向上	YOYO
前锋	0.2677	0.2022	0.1338	0.1338	0.0603	0.2022
中场	0.2401	0.1974	0.0995	0.0995	0.0612	0.3023
后卫	0.2095	0.1783	0.0989	0.0989	0.1166	0.2976
守门员	0.1895	0.2150	0.1218	0.1218	0.1624	0.1895

5.3.2 客观权重计算模型的建立与求解

采用熵权法（EWM）确定各种项目的客观权重。

步骤一：列出运动员得分矩阵 $Z(z_{ij})_{n \times m}$

n 为运动员的数目， m 为体能测试的项目数。

步骤二：计算概率矩阵 $P(p_{ij})_{n \times m}$

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^n z_{ij}} \tag{5}$$

步骤三：计算每个指标的信息熵

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (j = 1, 2, \dots, m) \tag{6}$$

步骤四：计算每个指标的信息效用值

$$d_j = 1 - e_j \tag{7}$$

步骤五：计算权重向量

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j} \quad (j = 1, 2, \dots, m) \tag{8}$$

表 5.11 客观权重求解结果

位置	30m 跑	箭头跑	立定跳远	纵跳	引体向上	YOYO
前锋	0.0686	0.1447	0.0873	0.0653	0.5129	0.1213
中场	0.0553	0.1317	0.1578	0.1311	0.3931	0.1310
后卫	0.0514	0.1850	0.0891	0.1190	0.3835	0.1719
守门员	0.0463	0.1669	0.0314	0.0762	0.2935	0.3856

5.3.3 组合计算模型的建立与求解

组合权重:在熵权法中,当重要性指标差异较小、非重要性指标差异较大时,可能出现重要性指标权重较小而非重要性指标权重较大的不合理现象。

为减少这种情况的发生,本文先用层次分析法进行赋值,保证重要性指标的权重,再采用熵权法计算出的信息效用值 d_j 对指标权重进行调节。这样既保证重要性指标所占权重,又中和了权重的主观性和客观性,得到更精确合理的权重。

$$w_j = \lambda \times w_j^S + (1 - \lambda) \times w_j^O, j = 1, 2, \dots, m \quad (9)$$

λ 为差异系数,计算方式为:

$$\lambda = \frac{m}{m-1} \left[\frac{2}{m} \sum_{j=1}^m p_j - \frac{m+1}{m} \right] \quad (10)$$

其中 p_j 是主观权重 w_j^S 按升序排列后的序列。

表 5.12 组合权重求解结果

位置	30m 跑	箭头跑	立定跳远	纵跳	引体向上	YOYO
前锋	0.22	0.19	0.12	0.12	0.17	0.18
中场	0.18	0.17	0.12	0.11	0.18	0.25
后卫	0.17	0.18	0.10	0.10	0.19	0.26
守门员	0.17	0.21	0.11	0.12	0.18	0.21

5.4 问题四模型的建立与求解

5.4.1 体能测试标准制定模型的建立与求解

首先,对于第 j 个测试指标,将所有测试球员的测试值记录在集合 Y_j 中:

$$Y_j = \{y_{ij} | i = 1, 2, \dots, n\}, j = 1, 2, \dots, m \quad (6)$$

接着,计算第 j 个指标的最小值 a_j 和最大值 b_j :

$$a_j = \min Y_j = \min_{1 \leq i \leq n} \{y_{ij}\} \quad (7)$$

$$b_j = \max Y_j = \max_{1 \leq i \leq n} \{y_{ij}\} \quad (8)$$

然后,将区间 $[a_j, b_j]$ 分成 100 等份(100 分制)或者 10 等份(10 分制),这里以 100 分制为例:

$$\Delta = \frac{b_j - a_j}{99} \quad (9)$$

$$x_k = a_j + k \times \Delta, k = 0, 1, 2, \dots, 99$$

这样, $x_j^0, x_j^1, x_j^2, \dots, x_j^{99}$ 就把区间 $[a_j, b_j]$ 分成了 99 等份

统计某个项目成绩在各个长度为 Δ 的小区间内的人数 z_i^k ，就可以得到该项目成绩的频数分布。

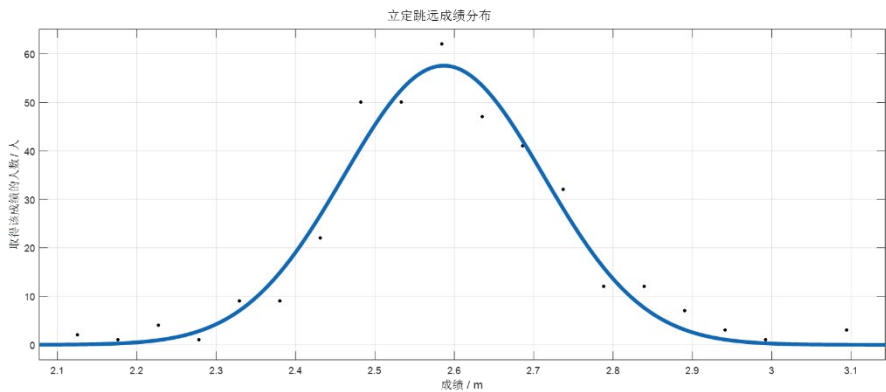
$$\tilde{x}_j^k = \frac{x_j^k + x_j^{k+1}}{2}, k = 0, 1, \dots, 98 \tag{10}$$

根据一系列的点 (\tilde{x}_j^k, z_j^k) ，就可以拟合出某个项目中得到某一成绩的人数关于成绩的概率密度函数，记为 $g_j(x)$ ，据此我们可以计算出该项目成绩的各级分位数：

$$P(X \leq \mu_j^p) = \int_{-\infty}^{\mu_j^p} g_j(x) dx = p \tag{10}$$

高斯分布拟合结果如下：

图 5.3 立定跳远成绩拟合图



$$f(x) = a1 * \exp(-((x - b1)/c1)^2)$$

表 5.12 体能素质指标拟合参数

	a1	b1	c1	SSE	R-square	Adjusted R-square	RMSE
立定跳远	57.58	2.587	0.1776	310.8	0.9612	0.9566	4.276
纵跳	36.78	54.26	14.25	2213	0.5668	0.5158	11.41
引体向上	25.42	14.56	9.976	3804	0.1532	0.05353	14.96
箭头跑	68.69	8.059	0.5832	562.3	0.9529	0.9474	5.751
30m	37.62	4.063	0.2146	248.30	0.9249	0.916	3.822
YOYO	35.18	891.4	348.7	3853	0.4677	0.4051	15.06

由图和表中数据，可见拟合结果较好。
根据拟合结果和求出的 100 个分位数，即可求的新的百分制评分标准，立定跳远成绩表如下图所示，其余成绩表详见附件。

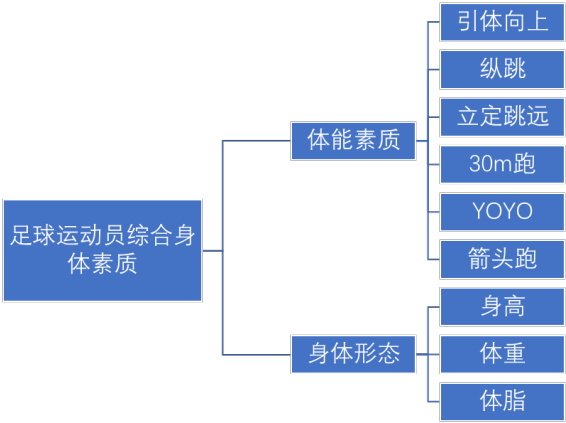
表 5.13 立定跳远成绩表

成绩	得分	成绩	得分	成绩	得分	成绩	得分
2.295	1	2.506	26	2.59	51	2.676	76
2.329	2	2.51	27	2.593	52	2.68	77
2.351	3	2.514	28	2.596	53	2.684	78
2.367	4	2.518	29	2.6	54	2.688	79
2.38	5	2.521	30	2.603	55	2.693	80
2.392	6	2.525	31	2.606	56	2.697	81
2.402	7	2.528	32	2.609	57	2.702	82
2.411	8	2.532	33	2.612	58	2.707	83
2.419	9	2.535	34	2.616	59	2.712	84
2.426	10	2.539	35	2.619	60	2.717	85
2.433	11	2.542	36	2.622	61	2.723	86
2.439	12	2.545	37	2.625	62	2.728	87
2.446	13	2.549	38	2.629	63	2.735	88
2.451	14	2.552	39	2.632	64	2.741	89
2.457	15	2.555	40	2.635	65	2.748	90
2.462	16	2.558	41	2.639	66	2.755	91
2.467	17	2.562	42	2.642	67	2.763	92
2.472	18	2.565	43	2.646	68	2.772	93
2.477	19	2.568	44	2.649	69	2.782	94
2.481	20	2.571	45	2.653	70	2.794	95
2.486	21	2.574	46	2.656	71	2.807	96
2.49	22	2.578	47	2.66	72	2.823	97
2.494	23	2.581	48	2.664	73	2.845	98
2.498	24	2.584	49	2.668	74	2.879	99
2.502	25	2.587	50	2.672	75	2.931	100

5.4.2 对问题三中组合计算模型的完善

除体能素质（耐力、爆发力）外，身体形态（如身高、体重等）对足球运动员的综合身体素质也有很大影响。通过阅读参考文献并研究相关测试标准，我们决定以身高、体重和体脂三个具有代表性的指标来衡量足球运动员的身体形态。在原有测试方法的基础上，由于各项测试与速度、上肢爆发力等体能素质因子的对应关系较为简单，故删去体能素质因子层，将测试与体能素质直接对应。建立新的指标层次结构如图所示：

图 4 加入身体形态后的指标层次结构



通过层次分析法，计算出 m 个指标的权重为 $w_j, j = 1, 2, \dots, m$ 。
一致性检验的结果和权重的计算结果如下，计算方法与第三问中相同：

表 5.14 身体形态权重表

身体形态因子	守门员			后卫		
	身高	体重	体脂	身高	体重	体脂
身高	1	3/2	2	1	1	2
体重	2/3	1	1	1	1	1
体脂	1/2	1	1	1/2	1	1

身体形态因子	中场			前锋		
	身高	体重	体脂	身高	体重	体脂
身高	1	1	3/2	1	1	3/2
体重	1	1	4/3	1	1	5/4
体脂	2/3	3/4	1	2/3	4/3	1

表 5.15 身体形态一致性检验数据

	体能素质		身体形态	
	CI	CR	CI	CR
守门员	0	0	0.0046	0.0089
后卫	0.068	0.06	0.0268	0.0516
前锋	0.0033	0.003	0.0018	0.0036
中场	-0.0168	-0.015	0	0.0015

表 5.16 各位置身体形态权重表

	中场	后场	守门员	前锋
身高	0.3767	0.4126	0.4638	0.3774
体重	0.3622	0.3275	0.2809	0.3552
体脂	0.2612	0.2599	0.2552	0.2674

通过计算出的比例即可将身高、体重和体脂与第三问中体能素质测试成绩统一运动员综合身体素质评估中。

然而，受数据所限，暂时无法将身体形态指标映射到具体成绩中。考虑到熵权法计算的权重完全由数据决定，具有极大的不稳定性，为了制定一个具有普适性和稳定性的指标，我们在此并没有采用熵权法和层次分析法结合的组合权重，具体分析详见模型评价、改进与推广。

5.4.3 球员体能成长性评估模型的建立与求解

通过收集足量球员的年龄与各项体能测试分数，我们可以计算出特定微小年龄段中球员各项身体素质的平均水平。从理论上说，由于人本身身体素质发展的规律性，球员各项身体素质的平均水平也应该符合统计规律，可通过回归分析，拟合出各项身体素质的平均水平同年龄的函数关系。进而，我们根据这个函数关系，就可以计算出一般意义下，运动员在某一年龄时的身体素质水平占其生涯最高水平的理论比例。

我们假设对于某个特定的球员，他在某一年龄时的身体素质水平占他所能达到的巅峰水平的比例等于先前所述的理论比例，那么在有了通过体能测试得到的当前身体素质水平后，就可以估计出未来他能达到的最高水平。结合年龄、当前身体素质水平、未来巅峰身体素质水平这三个数据，我们就可以对一个球员的成长性进行评估。

记各项身体素质指标的平均水平同年龄的函数关系为：

$$\hat{y}_j = f_j(x), j = 1, 2, \dots, m \tag{1}$$

y_i 指第 j 个指标的总体平均分数， x 指年龄

我们收集了 2017 年全年活跃足球运动员数据集（包含近 20000 名足球运动员的速度、耐力、爆发力等 50 多项指标，具体数据集见附件），对其进行预处理并多项式拟合，我们得到了足球运动员的各项体能素质与年龄的关系 \hat{y}_j ，拟合结果如图：

图 5.5 守门员年龄与下肢爆发力的关系

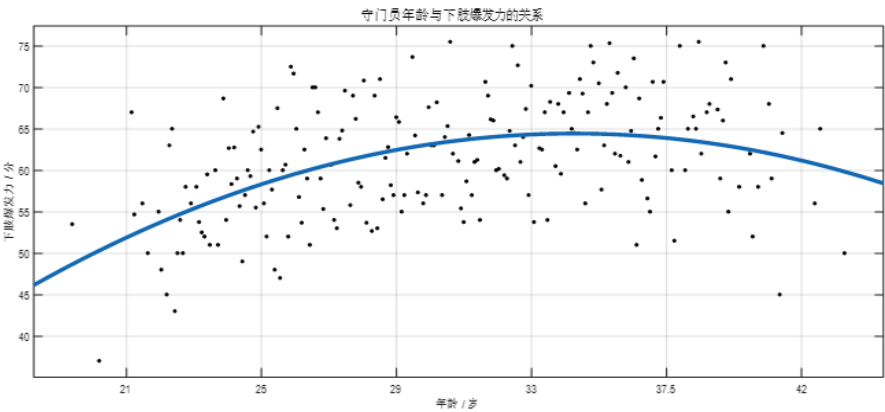


表 5.16 体能素质拟合参数
 $f(x) = p1 * x^2 + p2 * x + p3$

	p1	p2	p3	SSE	R-square	Adjusted R-square	RMS E
上肢爆发	-0.00038	0.2967	8.038	6917	0.1086	0.09971	5.881
速度	-0.00041	0.3016	-11.19	6577	0.08866	0.07977	5.664
下肢爆发	-0.00046	0.3802	-14.55	8071	0.2264	0.2187	6.352
耐力	-0.00027	0.1993	-3.273	5035	0.06231	0.05329	4.92
敏捷	-0.00063	0.4918	-49.68	11500	0.1604	0.1524	7.41

由图和表中数据，可见函数拟合情况较为良好，由此可得到综合身体素质指标平均水平函数为：

$$\hat{y} = f(x) = \sum_{j=1}^m w_j \hat{y}_j = \sum_{j=1}^m w_j f_j(x) \tag{2}$$

综合身体素质指数平均水平函数的最大值为 \hat{y}^{max}
如果一个球员年龄为 x_i ，第 j 项体能测试测得的成绩为 y_{ij} ，则他目前的综合身体素质指数为

$$y_i = \sum_{j=1}^m w_j y_{ij} \tag{3}$$

记他未来的巅峰身体素质指数为 y^{max} ，则根据假设有：

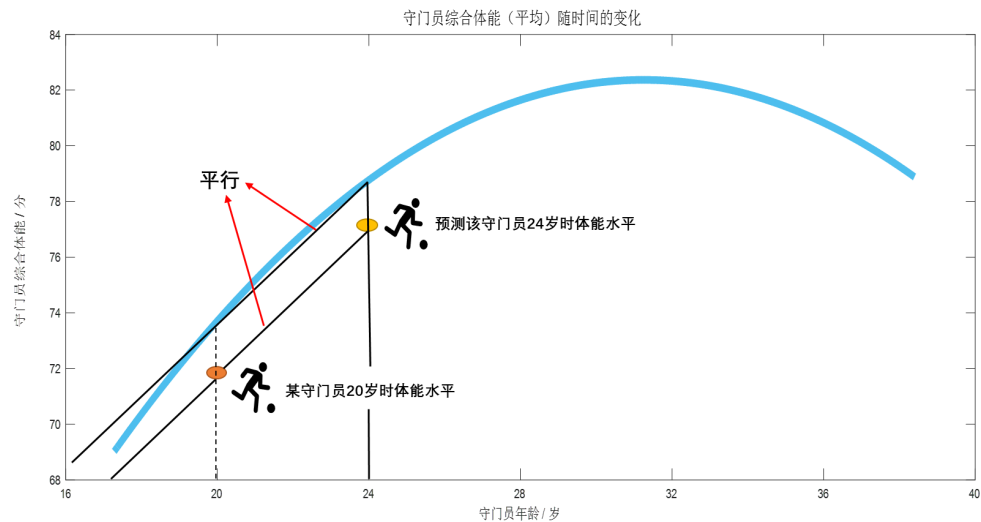
$$\frac{y_i}{y_i^{max}} = \frac{f(x_i)}{\hat{y}^{max}} \tag{4}$$

即：

$$y_i^{max} = \frac{y_i \times \hat{y}^{max}}{f(x_i)} \tag{5}$$

这样，我们就可以通过数对 (y_i, y_i^{max}) 衡量该球员的当下综合身体素质指数及其成长空间。如下图所示：

图 5.7 计算示意图



六、模型的评价、改进与推广

6.1 模型的优点与缺点

体能测试项目权重确定模型

优点：该模型结合了主观的层次分析法和客观的熵权法，并采用了有序加权平均算子对主客观权重进行加权，既吸纳了主观意见，又充分利用了客观数据，达到了主客观结合的效果。对于不同位置的球员，可以计算出具有区分度的权重，从而更合理地评估球员的体能素质水平。

缺点：由于熵权法确定的权重是由数据计算得出的，该客观权重会随着数据的变化而变化，稳定性较差，因此不适合作为一组普适性的权重；可只取用层次分析法所确定的主观权重作为最终权重。此外，第四问中新加入的身体形态指标，由于未能得到相应标准，只能求出各指标权重，无法求出各指标所对应成绩。

2) 球员成长性评估模型

优点：除了原有的体能素质指标外，该模型加入了身体形态指标，从而对球员的身体素质有一个较为全面的评估。另外，该模型也考虑了球员的年龄的影响。对于球员成长空间的评估是基于大量权威的职业运动员各项体能素质对时间的拟合数据来进行的，从理论上来说是较为合理的。

缺点：由于题目原有的数据中缺少年龄信息，故我们无法用我们的模型计算出这些球员的成长性指数，该模型的实际应用效果还有待检验。

3) 体能测试标准制定模型

优点：该模型基于对题目原有数据的统计与分析，拟合出的概率密度函数与高斯分布符合得很好，这说明了模型的合理性。

缺点：由于已有的数据还不够多，拟合出的函数还不够精确，这影响了最终制定出的标准的准确性。

6.2 模型的改进与推广

1) 体能测试项目权重确定模型

由于熵权法确定的权重是由数据计算得出的，该客观权重会随着数据的变化而变化，因此在大多数情况下，它不适合作为一组普适性的权重；只有当收集的数据足够大，足以代表整体的时候，熵权法确定的权重才可能不会因为特定情境下收集到特定的数据而发生较大的变化，但这时确定出的各测试项目的权重可能相差不大。所以，如果要计算出一组具有普适性、稳定性的权重，建议采用基于专家打分比较确定的主观权重，是否采用客观权重视数据量的大小及客观权重的差异性而定。

2) 球员成长性评估模型

该成长性模型同样可用于其他体育运动中运动员的成长性评估。

3) 体能测试标准制定模型

该体能测试标准制定模型同样可用于其他体能测试的标准制定中。

七、参考文献

[1] U17 足球运动员专项灵敏和耐力素质分析及评价等级的建立_陈翀
[2] 青少年足球运动员位置体能特征研究_何光丰
[3] 现代足球训练体系下的运动员体能测试方法应用分析_陈旭
[4] 足球天才识别与发展理论模型与应用体系研究_梁斌

附录

附录 1
介绍：附件清单
1. matlab 代码 2. python 代码 3. 各项中间数据及最终处理数据 4. 2017 全球活跃足球运动员数据集 5. 层次分析调查问卷

附录 2

介绍：基于熵权法计算守门员权重

```

import math
import xlrd
file_location = "D:\\临时文件\\数学建模\\2021 校赛\\数据及处理结果\\守门员.xls"
data = xlrd.open_workbook(file_location)
sheet = data.sheet_by_index(0)
Z=[[sheet.cell_value(i,j) for i in range(sheet.nrows)]for j in range(1,7) ]
#计算概率矩阵,信息熵,熵权
p=[]
for i in Z:
    p.append([j/sum(i) for j in i])
n=sheet.nrows
E=[]
def ln(x):
    if x==0:
        return 0
    elif x>0:
        return math.log(x)
for i in p:
    e=(-1)*sum([j*(ln(j)) for j in i])/ln(n)
    E.append(e)
W=[round((1-e)/sum([1-e for e in E]),4) for e in E]
print('守门员权重: ',W)

```

附录 3

介绍：其他各项利用体能测试标准制定模型制定出的新的标准。

YOYO 成绩表 成绩单位：米

成绩	得分	成绩	得分	成绩	得分	成绩	得分
318	1	733	26	898	51	1066	76
385	2	740	27	904	52	1074	77
428	3	748	28	910	53	1082	78
460	4	755	29	916	54	1090	79
486	5	762	30	922	55	1099	80
508	6	769	31	929	56	1108	81
528	7	776	32	935	57	1117	82
545	8	783	33	941	58	1127	83
561	9	790	34	948	59	1137	84
575	10	796	35	954	60	1147	85

589	11	803	36	960	61	1158	86
602	12	810	37	967	62	1169	87
614	13	816	38	973	63	1181	88
625	14	823	39	980	64	1194	89
636	15	829	40	986	65	1207	90
646	16	835	41	993	66	1222	91
656	17	842	42	1000	67	1238	92
666	18	848	43	1007	68	1255	93
675	19	854	44	1014	69	1275	94
684	20	860	45	1021	70	1297	95
693	21	867	46	1028	71	1323	96
701	22	873	47	1035	72	1355	97
709	23	879	48	1043	73	1398	98
717	24	885	49	1050	74	1465	99
725	25	891	50	1058	75	1579	100
纵跳成绩表 成绩单位：厘米							
成绩	得分	成绩	得分	成绩	得分	成绩	得分
30.82	1	47.78	26	54.51	51	61.38	76
33.57	2	48.09	27	54.77	52	61.7	77
35.31	3	48.39	28	55.02	53	62.04	78
36.62	4	48.68	29	55.27	54	62.39	79
37.69	5	48.98	30	55.53	55	62.74	80
38.59	6	49.26	31	55.78	56	63.11	81
39.39	7	49.55	32	56.04	57	63.48	82
40.1	8	49.83	33	56.29	58	63.87	83
40.75	9	50.1	34	56.55	59	64.28	84
41.35	10	50.38	35	56.81	60	64.7	85
41.9	11	50.65	36	57.07	61	65.15	86
42.42	12	50.92	37	57.34	62	65.61	87
42.91	13	51.18	38	57.6	63	66.1	88
43.37	14	51.45	39	57.87	64	66.62	89
43.82	15	51.71	40	58.14	65	67.17	90
44.24	16	51.97	41	58.42	66	67.77	91
44.65	17	52.23	42	58.69	67	68.42	92
45.04	18	52.48	43	58.97	68	69.13	93
45.41	19	52.74	44	59.26	69	69.93	94
45.78	20	52.99	45	59.54	70	70.83	95

46.13	21	53.25	46	59.84	71	71.9	96			
46.48	22	53.5	47	60.13	72	73.21	97			
46.82	23	53.75	48	60.43	73	74.95	98			
47.14	24	54.01	49	60.74	74	77.7	99			
47.46	25	54.26	50	61.06	75	78.71	100			
箭头跑成绩表 成绩单位：秒										
成绩	得分	成绩	得分	成绩	得分	成绩	得分			
7.1	1	7.79	26	8.07	51	8.35	76			
7.21	2	7.81	27	8.08	52	8.36	77			
7.28	3	7.82	28	8.09	53	8.38	78			
7.34	4	7.83	29	8.1	54	8.39	79			
7.38	5	7.84	30	8.11	55	8.41	80			
7.42	6	7.85	31	8.12	56	8.42	81			
7.45	7	7.87	32	8.13	57	8.44	82			
7.48	8	7.88	33	8.14	58	8.45	83			
7.51	9	7.89	34	8.15	59	8.47	84			
7.53	10	7.9	35	8.16	60	8.49	85			
7.55	11	7.91	36	8.17	61	8.5	86			
7.57	12	7.92	37	8.18	62	8.52	87			
7.59	13	7.93	38	8.2	63	8.54	88			
7.61	14	7.94	39	8.21	64	8.56	89			
7.63	15	7.95	40	8.22	65	8.59	90			
7.65	16	7.97	41	8.23	66	8.61	91			
7.67	17	7.98	42	8.24	67	8.64	92			
7.68	18	7.99	43	8.25	68	8.67	93			
7.7	19	8	44	8.26	69	8.7	94			
7.71	20	8.01	45	8.28	70	8.74	95			
7.73	21	8.02	46	8.29	71	8.78	96			
7.74	22	8.03	47	8.3	72	8.83	97			
7.75	23	8.04	48	8.31	73	8.91	98			
7.77	24	8.05	49	8.32	74	9.02	99			
7.78	25	8.06	50	8.34	75	9.18	100			
引体向上成绩表 成绩单位：个										
成绩	6	9	11	13	15	16	18	20	23	27
得分	1	2	3	4	5	6	7	8	9	10
30m 跑成绩表 成绩单位：秒										
成绩	得分	成绩	得分	成绩	得分	成绩	得分			
3.71	1	3.965	26	4.067	51	4.17	76			

3.751	2	3.97	27	4.071	52	4.175	77
3.778	3	3.975	28	4.074	53	4.18	78
3.797	4	3.979	29	4.078	54	4.185	79
3.813	5	3.983	30	4.082	55	4.191	80
3.827	6	3.988	31	4.086	56	4.196	81
3.839	7	3.992	32	4.09	57	4.202	82
3.85	8	3.996	33	4.094	58	4.208	83
3.86	9	4	34	4.098	59	4.214	84
3.869	10	4.005	35	4.101	60	4.22	85
3.877	11	4.009	36	4.105	61	4.227	86
3.885	12	4.013	37	4.109	62	4.234	87
3.892	13	4.017	38	4.113	63	4.241	88
3.899	14	4.021	39	4.117	64	4.249	89
3.906	15	4.025	40	4.121	65	4.257	90
3.912	16	4.028	41	4.126	66	4.266	91
3.918	17	4.032	42	4.13	67	4.276	92
3.924	18	4.036	43	4.134	68	4.287	93
3.93	19	4.04	44	4.138	69	4.299	94
3.935	20	4.044	45	4.143	70	4.313	95
3.941	21	4.048	46	4.147	71	4.329	96
3.946	22	4.052	47	4.151	72	4.348	97
3.951	23	4.055	48	4.156	73	4.375	98
3.956	24	4.059	49	4.161	74	4.416	99
3.961	25	4.063	50	4.165	75	4.478	100