

第六届湖南省研究生数学建模竞赛

题 目 黄桃品质分析及贮藏时间预测

摘 要：

黄桃是商用价值较高的一款水果产品。为避免黄桃腐烂，需要精准地把握采摘时机，并通过合适的贮藏、运输方式，以保持黄桃的最佳风味。实际中，黄桃品质的评判方法分为机械指标评分和人工评分。为了更加科学地评判黄桃品质，需要综合利用这两种评判方法。本文通过建立回归模型分析机械指标与人工评分之间的关系，并构建了预测黄桃品质的定量模型。另外，通过分析贮藏温度和时长对黄桃品质的影响，构建了黄桃品质随贮藏温度和贮藏时间变化的预测模型。该模型可以根据黄桃的初始品质为商家推荐合适的贮藏温度和贮藏时长。

针对问题 1，通过线性回归、逐步线性回归方法，根据 2018 年至 2020 年的实验数据，分析了黄桃样本的各项机械指标评分与人工评分之间的关系以及两类方法的评判一致性，并构建黄桃品质分析的定量模型。

针对问题 2，基于问题 1 构建的黄桃品质分析定量模型，根据 2020 年度的实验数据建立模型；并利用该模型判断附件 7 中“样本 A”至“样本 I”9 个测试样本的人工评分等级（“优质”或“非优质”）。

针对问题 3，考虑到黄桃品质的影响因素不仅与贮藏环境和贮藏时间有关，还与黄桃采摘后的初始指标有关。单纯的考虑贮藏环境和贮藏时间的影响难以准确预测黄桃品质的变化规律，因此需要综合考虑包括各项机械指标在内的多方面因素对于黄桃品质的影响。为此，我们考虑将同一批次的机械指标值作为时序数据，利用其训练循环神经网络模型（LSTM）。模型可以根据各项机械指标的初始值预测其变化趋势，从而选择最利于果农的贮藏方式和环境。

关键词：逐步线性回归；LSTMCell；定量模型；预测；黄桃品质

1. 问题重述

1.1 问题背景

黄桃是分布于我国长江流域及北部地区的早熟大果型桃,一般于6月中下旬成熟,果面淡黄至金黄,果肉金黄,果汁丰富。果实成熟后可挂树10天左右,采摘后对比同期其他品种的储运能力也较强,是商用价值较高的一款水果产品。然而,黄桃比较容易腐烂。当黄桃从树上采摘下来后,需要及时贮藏以保持口感风味,而不同的存储温度带来了不同的运输和贮藏成本,并可能影响黄桃品质。目前普遍认为在冷库(1°C)贮藏情况下,低温会影响黄桃代谢,风味有所下降,但保质期较长;在冷柜(8°C)贮藏情况下,黄桃代谢情况较好,但保鲜期较短。因此,需要精准把握其采摘时机以及贮藏、运输方式,才能在风味最佳时上架各个水果店铺。

黄桃品质与很多因素有关,比如可溶性固形物(TSS)、硬度、色差等。这些因素统称为机械指标,共同决定了黄桃品质的优劣。对机械指标的测定属于有损测试,如利用FTA硬度计、SMS质构仪对黄桃质地(或硬度)进行有损测试、利用手持式折光仪对黄桃TSS含量进行有损测试等。根据专家经验,优质黄桃的机械指标评分应以果实硬度 $>1.5\text{kg}/\text{cm}^2$, TSS含量下降率 $<20\%$ (即TSS初始指标和贮藏期间的指标对比)来度量。而人工评分时,当色泽、质地和味道这三个指标均分 >80 分时,则判定为“优质”黄桃。目前采用双评分机制,普遍认知是机械指标比较容易满足,而人工评分要求更高。当前对果实品质多以机械指标评判,而感官评价也代表了果实的综合品质和性状。

1.2 文献综述

市场对高品质果品的需求促进了研究者对影响水果品质的许多因素进行了研究。林震山等以黄桃的感观评价、呼吸强度、TSS、维生素含量、失重率、可滴定酸含量和硬度为考核指标,探讨了贮藏温度与黄桃贮藏品质的关系^[1]。结果表明:在 $-1^{\circ}\text{C}\sim 2^{\circ}\text{C}$ 下贮藏的黄桃能够保持良好色泽和较高硬度,营养成分保持得较好;在 $-1^{\circ}\text{C}\pm 0.5^{\circ}\text{C}$ 下贮藏35天时,黄桃品质仍保持较高水平。蒋均等以连平鹰嘴蜜桃为材料,探究了不同贮藏温度与包装对贮藏效果的影响^[2]。余意等研究了果实采收成熟度和贮藏温度对黄桃完熟品质的影响^[3]。吴剑等研究发现,通过合适的热处理-冷藏方法,可以改善贮藏后的黄桃品质^[4]。

在建模和预测问题中,通常需要研究两个或多个变量之间的关系,即进行一元或多元线性回归分析。一元线性回归分析的主要任务是以一个变量估计另一个变量,若两个变量之间的数量变化呈近似线性关系,则可以建立一元线性回归方程。在多元线性回归分析中,逐步回归分析是一种被广泛应用于多个学科领域的自变量选择方法,其基本思想在于逐个引入显著性变量,同时剔除不显著性变量,以建立最优或合适的回归模型,从而更加深入地研究变量之间的依赖关系^[5]。通过逐步回归分析方法,许鑫研究了不同品种绵羊体重与体尺指标之间的相关关系^[6];吴巧英筛选出对缎纹试样、平纹试样和全体试样经向缝纫平整度主客观评价起关键影响作用的结构及性能指标^[7];陆会宁等研究了兰小杂交F1代羔羊的体重与体尺之间的关系,得到了体重与体尺的最优回归方程^[8];李松臣和张世英分析了收入水平、人口组成、教育水平、物价水平等因素对人口出生率的影响,得出人口组成和收入水平是人口出生率的重要影响因素,并给出了相关的线性回归

模型^[9]；王红等对进出口贸易预测进行了建模^[10]。

机器学习领域的神经网络学习也被广泛用于解决预测问题，如循环神经网络（RNN）、卷积神经网络等^[11]。其中，RNN 多被用于时间序列预测，即利用过去一段时间内某事件的特征预测未来该事件的特征。常见的循环神经网络包括长短期记忆（LSTM）神经网络，其相比简单 RNN 更适于解决较长时间序列预测问题，因而被广泛应用于多个领域。借助 LSTM 神经网络，王鑫等对系统级故障时间序列数据进行了可靠预测^[12]；彭燕等预测了股票价格走势^[13]；王渊明使用某电商公司历史销售数据预测了未来销量^[14]。

鉴于回归分析和 LSTM 神经网络在预测和建模问题上的广泛应用并能获得可观效果，本研究利用它们对黄桃品质的影响因素进行分析和对黄桃贮藏品质进行预测。

1.3 问题提出

位于 S 市的某一黄桃种植基地通过多种实验手段对基地内的 X 品牌黄桃进行了品质评测，其实验思路主要是在黄桃成熟期的一周左右分三批进行采摘，并设置冷库（1℃）和冷柜（8℃）两种储存方式，每隔 3 至 5 天抽取出一部分样本测量各项机械指标与相对应的人工品尝评分。为帮助农户与商户更好的采摘、存运与售卖，推动 S 市黄桃产业的数字化转型，需要通过数据分析来刻画黄桃品质与贮藏方式之间的关系。

如何利用机械指标和人工评分，更加科学的评判果实品质，是本研究的主要目的。结合实际，根据附件 1~附件 7 中分别提供的 2018 至 2019 年黄桃各项机械指标和人工打分数据，通过建立数学模型研究贮藏方式对黄桃品质的影响，主要解决以下问题：

（1）根据附录 1 中的实验描述和附件 1~附件 7 中的实验数据，分析机械指标、人工评分之间的关系与两类方法的评判一致性，构建黄桃品质分析的定量模型，为商家推荐优质黄桃；比较模型得到的结论与“专家经验”以及“普遍认知”之间的差异。

（2）2020 年度的评测中取消了人工评分，按照每批次采摘的 3 个测试样本的方式，在冷柜贮藏方式下随机抽出共 9 个样本，分别对应附件 7 中的“样本 A”至“样本 I”。结合问题 1 中的模型和结论，根据 2020 年度的数据，建立模型判断上述 9 个样本的人工评分结果是否为“优质”。

（3）根据附录 1 中的实验描述以及附件 1~附件 7 中的实验数据，综合分析贮藏环境（冷库或冷柜）和贮藏时长对黄桃品质的影响，并构建黄桃品质随贮藏环境和贮藏时间变化的预测模型。通过预测模型，能够根据黄桃的初始品质为商家推荐合适的贮藏方式和时间，同时将预测结果与“普遍认为”进行比较。若下一年度果农需要用该模型预测采摘黄桃的贮藏期，果农需要进行多少次有损测试。

2. 回归分析与模型建立

2.1 机械指标、人工评分之间的关系

本研究拟使用回归分析方法研究黄桃各项机械指标与人工评分之间的关系。在进行回归分析之前，需要对实验数据进行正态性检验。对于 2018 年和 2019 年人工评分数据，K-S 检验结果显示其服从正态分布，检验结果见附录。基于这一检验结果，本研究主要使用线性回归方法研究了机械指标与人工评分的相关性。

2.1.1 色泽的机械指标与人工评分之间的关系

参考附件 2，色泽的机械指标由国际照明委员会建立的颜色体系表征，分为 L、a、b、c、h 五个颜色维度。我们将 2018 年黄桃色泽的机械测量值 L、a、b、c、h 作为自变量，人工色泽评分作为因变量，运用逐步回归分析方法，得到 2018 年的五个颜色维度分别与人工评分无关。同样地，我们将 2019 年黄桃果皮的 L_1 、 a_1 、 b_1 、 c_1 、 h_1 ，及果肉的 L_2 、 a_2 、 b_2 、 c_2 、 h_2 作为输入自变量，人工色泽评分作为输出因变量，通过逐步回归分析方法将 L_1 、 b_1 、 c_1 、 h_1 、 L_2 、 a_2 、 b_2 、 c_2 、 h_2 这 9 个变量排除，仅剩下 a_1 作为自变量输入。最后，我们将 2019 年的黄桃果皮的 a 值作为自变量 x_1 、人工色泽评分作为因变量 y_1 ，进行线性回归建模，得到结果：

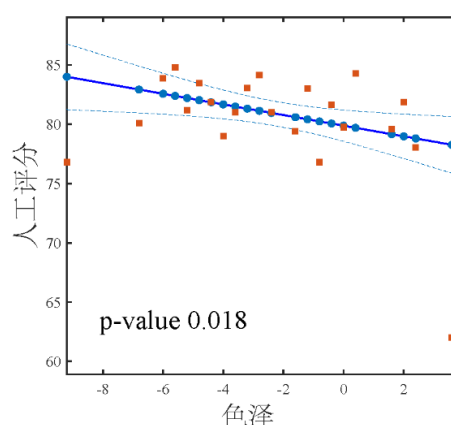


图 2.1 2019 年果皮色泽机械指标评分与人工评分之间的关系

线性回归模型如下：

表 2.1 果皮色泽机械指标评分与人工评分线性回归结果

2019 色泽	系数估计	标准误差	t 统计量	p 值
截距 β_1	79.884	0.65609	121.76	1.4107e-67
x_1	-0.45249	0.18529	-2.4421	0.017908
线性回归方程		$y_1 = -0.45249x_1 + 79.884$		

2.1.2 色差的机械指标与人工评分之间的关系

基于对黄桃色泽的描述，可以用 (L, a, b) 或 (L, c, h) 表示一个特定的颜色。由于前一种简单易懂，行业应用很普遍，我们在计算色差时采用 (L, a, b) 体系。在该体系中，如果用三维立体直角坐标系来定位一种颜色，则 L 表示颜色深浅坐标，a 表示红绿轴坐标，b 表示黄蓝轴坐标。于是空间中的任意一种颜色可以用 (L^*, a^*, b^*) 表示。

两种颜色 (L_1, a_1, b_1) 和 (L_2, a_2, b_2) 的色差可以用坐标差来表示：

$$\Delta E = \frac{(\Delta L^2 + \Delta a^2 + \Delta b^2)}{2} \quad (2.1)$$

其中， $\Delta L = |L_1 - L_2|$ ， $\Delta a = |a_1 - a_2|$ ， $\Delta b = |b_1 - b_2|$ 。

接下来，分析色泽的机械指标与人工评分之间的关系。我们将 2018 年、2019

年数据分开分析。首先，对同一批次黄桃的色泽机械指标两两配对，并按公式(2.1)计算黄桃的色差，对应人工色泽评分两两之间的差别，进行线性回归分析，得到结果：

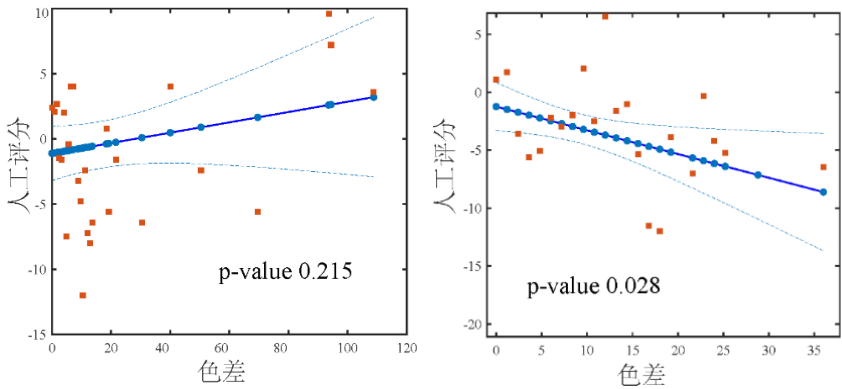


图 2.2 2018 年、2019 年黄桃色差机械指标评分与人工评分之间的关系

线性回归结果显示，2019 年黄桃样本色泽的机械指标差异与人工评分差异呈显著正相关关系(p 值 < 0.05)，而 2018 年黄桃样本中两者关系不显著。结果间接表明了 2019 年黄桃的色泽测定中机械指标与人工评分的一致性。

2.1.3 质地的机械指标与人工评分之间的关系

黄桃的硬度测量使用装有 8 mm 直径探针的 FTA-XT Plus 质地分析仪 (Stable Micro-Systems; Texture Technologies Corp., Scarsdale, NY) 测量果肉质 (硬度)。进行硬度测试时，探针的测试前速度为 1 mm/s，测试速度为 2 mm/s，距离为 5 mm。硬度值以千克表示 (单位 kg/cm^2)。对于测由黄桃的左、右两侧硬度的数据，取平均值记录。

将黄桃样本的硬度机械指标作为自变量 x_2 ，质地的人工评分作为因变量 y_2 ，进行线性回归建模，得到结果：

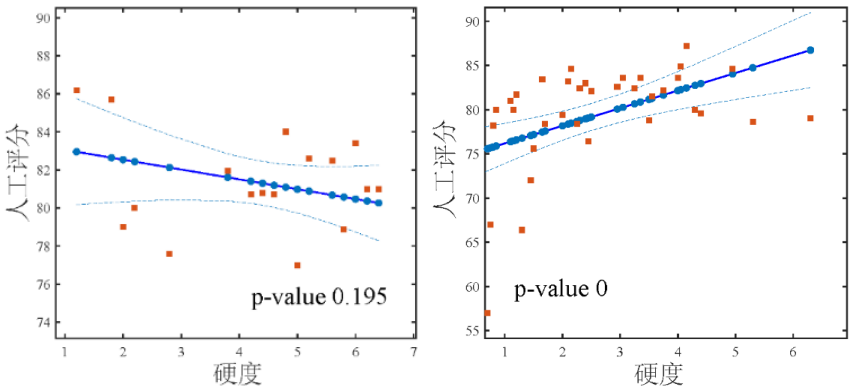


图 2.3 2018 年、2019 年黄桃硬度机械指标评分与人工评分之间的关系

线性回归结果显示，2019 年黄桃样本硬度机械指标与质地人工评分呈显著正相关关系，而 2018 年黄桃样本中两者关系不显著。所以，我们采用 2019 年黄桃样本建立硬度机械指标评分 x_2 与质地人工评分 y_2 的线性回归模型：

表 2.2 黄桃硬度机械指标评分与人工评分线性回归结果

2019 质地	系数估计	标准误差	t 统计量	p 值
截距 β_2	74.17	1.5667	47.342	2.5941e-41
x_2	1.9925	0.53018	3.7582	0.00047227
线性回归方程		$y_2 = 1.9925x_2 + 74.17$		

2.1.4 TSS 的机械指标与人工评分之间的关系

可溶固形物总量（TSS，Total Soluble Solid）：食品行业一个常用的技术参数，主要是指可溶性糖类，使用 α 手持折光仪（ATAGO-1; Atago Co. Ltd., 东京，日本）在 20℃ 下，用蒸馏水校准，从单个水果的浸没中果皮的未稀释倾析汁液中测定可溶固形物总量，指标单位%。

对于 2018 年、2019 年黄桃样本，将每一批黄桃样本的 TSS 均值作为自变量，相应批次的味（糖酸、可固）人工评分作为因变量，进行线性回归建模，得到结果：

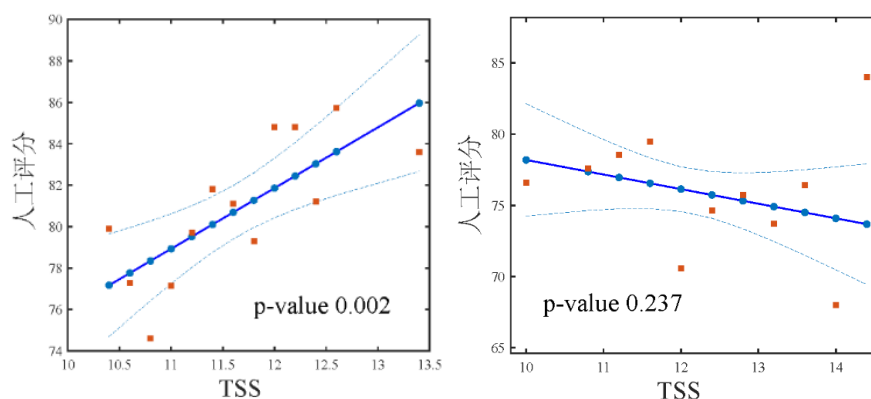


图 2.4 2018 年、2019 年黄桃 TSS 机械指标评分与人工评分之间的关系

线性回归结果显示，2018 年黄桃样本 TSS 机械指标与味（糖酸、可固）人工评分呈显著正相关关系，而 2019 年黄桃样本中两者关系不显著。

所以，我们采用 2018 年黄桃样本建立 TSS 机械指标 x_3 与味（糖酸、可固）人工评分 y_3 的线性回归模型：

表 2.3 黄桃 TSS 机械指标评分与人工评分线性回归结果

2018TSS	系数估计	标准误差	t 统计量	p 值
截距 β_3	47.494	9.861	4.8163	4.5823e-05
x_3	2.8589	0.84489	3.3837	0.00213
线性回归方程		$y_3 = 2.8589x_3 + 47.494$		

2.1.5 TSS 变化率的机械指标与人工评分之间的关系

针对每一批次黄桃样本，我们对贮藏期黄桃样本的 TSS 变化率进行计算，其中 TSS 变化率=（观测批次 TSS 均值-初始批次 TSS 均值）/ 初始批次 TSS 均值。

将 TSS 变化率作为自变量，味道（糖酸、可固）人工评分作为因变量，进行线性回归建模，结果如下：

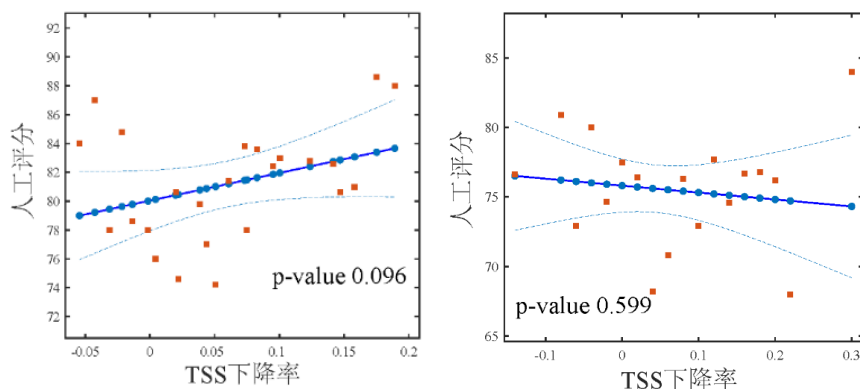


图 2.5 2018 年、2019 年黄桃 TSS 下降率机械指标评分与人工评分之间的关系

线性回归结果表明，黄桃的 TSS 下降率机械指标评分与味道（糖酸、可固）人工评分无显著相关性。

2.2 黄桃品质分析定量模型

根据上一节，我们得到关于黄桃色泽、质地和味道三个指标的线性回归模型，同时，我们将人工评分时，当色泽、质地和味道这三个指标均分 ≥ 80 分时认为样品“优质”。故，结合三个指标的线性回归模型，得到黄桃品质分析定量模型：

$$y = \begin{cases} 1, & \frac{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \beta_1 + \beta_2 + \beta_3}{3} \geq 80, \\ 0, & \text{其他.} \end{cases} \quad (2.2)$$

其中， $\alpha_1 = -0.45249$ ， $\alpha_2 = 1.9925$ ， $\alpha_3 = 2.8589$ ， $\beta_1 = 79.884$ ， $\beta_2 = 74.17$ ， $\beta_3 = 47.494$ ，分别对应三个线性回归模型中的系数。

根据专家经验，优质黄桃的机械指标评分应以果实硬度 $> 1.5\text{kg/cm}^2$ ，TSS 含量下降率 $< 20\%$ （即 TSS 初始指标和贮藏期间的指标对比）来度量。而人工评分时，当色泽、质地和味道这三个指标均分 ≥ 80 分时认为样品“优质”。对于目前的双评分机制，普遍认知是机械指标比较容易满足，而人工评分要求更高。

对比本节模型与上述提到的“专家经验”和“普遍认知”：

（1）TSS 含量下降率不在模型考虑范围内；

（2）如果我们将 80 分作为区分黄桃质地优质与否的界限，则根据黄桃质地人工评分的线性回归模型计算，优质黄桃的果实硬度 $\geq 2.9260\text{kg/cm}^2$ ；

（3）以黄桃质地为例，若按传统果实硬度评判指标，仅有 1 例为“非优质”；按本节模型，有 6 例为“非优质”，所以本文模型下的机械指标更为严苛。

2020 年度的评测中取消了人工评分，随机抽出冷柜贮藏方式下每批次采摘的 3 个测试样本（附件 7 中的“样本 A”-“样本 I”），模型判断这 9 个样本的人工评分结果如表 2.4 所示，9 个测试样本的三个指标预测评分如图 2.6 所示。

表 2.4 黄桃品质分析定量模型对人工评分等级的判断结果

样本	品质	样本	品质	样本	品质
样本 A	优质	样本 D	优质	样本 G	优质
样本 B	优质	样本 E	优质	样本 H	优质
样本 C	优质	样本 F	非优质	样本 I	优质

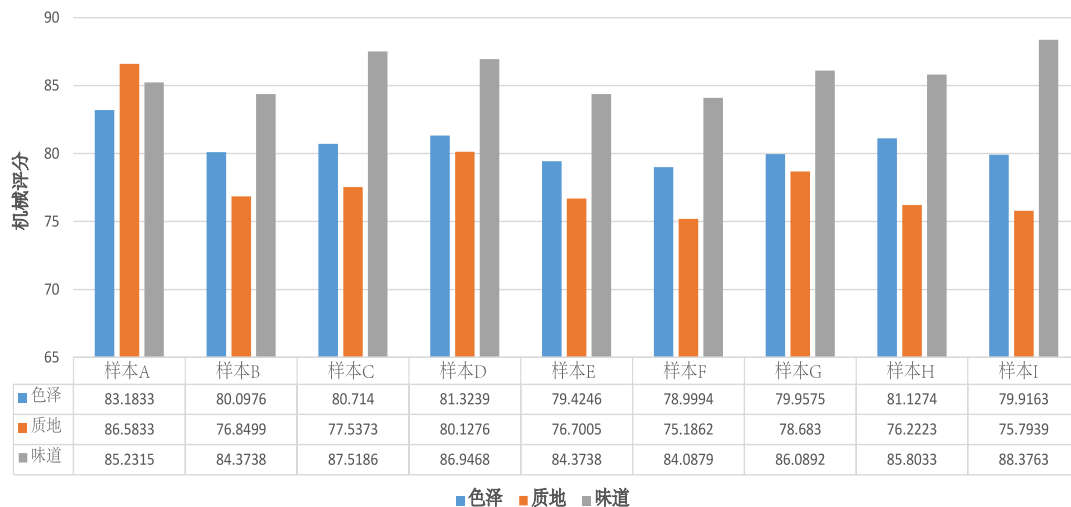


图 2.6 色泽、质地及味道三个指标的预测评分

3. 基于时间序列神经网络的预测模型

3.1 贮藏环境和时长的影响

由于不同的贮藏环境会影响黄桃保质期长短，随着贮藏时间的增加，冷库 1℃贮藏环境中的黄桃优质率明显高于 8 度贮藏环境中的黄桃优质率。研究表明，采摘后的贮藏环境对黄桃的保鲜有着重要的影响^[1]。低温可有效抑制黄桃的呼吸强度，延缓 VC 的降解，减少黄桃营养成分的消耗，降低失重率，有效保持黄桃的硬度和食用品质。综上，从给定数据和当前研究结果可知贮藏环境和贮藏时长对黄桃品质的影响如下：

- (1) 无论贮藏环境如何，黄桃品质会随着贮藏时长增加而下降。
- (2) 低温的贮藏环境能有效延长黄桃的保质期。
- (3) 稍高温度的贮藏环境不会影响采后初期黄桃品质，但随着贮藏时长增加。温度较高的贮藏环境会使得黄桃品质的下降速度快于温度较低的贮藏环境。

3.2 模型构建理念

黄桃品质的影响因素不仅与贮藏环境和贮藏时间有关，还与黄桃采摘后的初始指标有关，仅考虑贮藏环境和贮藏时间的影响难以准确预测黄桃品质的变化规律，因此需要综合考虑包括各项机械指标在内的多因素对于黄桃品质的影响。考虑到人工打分的方法会耗费大量的人力，而且带有一定的主观性，我们希望通过历史数据中学习出黄桃机械指标的变化规律，为黄桃贮藏方式和时间的选择提供更为科学的依据。

我们考虑将同一批次的机械指标值作为时序数据，利用其训练循环神经网络模型（LSTM）。根据黄桃机械指标的预测模型，可通过机械指标的预测值对

黄桃品质进行评估。所以，给定各项机械指标的初始值即可预测该指标的变化趋势，从而选择最有利于果农的贮藏方式和环境。

3.3 模型构建细节

我们按下列步骤构建黄桃机械指标值预测模型。

提取 2018 年和 2019 年的各项机械指标值，由于机械指标是由机器测定，故此处假设 2018 年的机械指标值有效，并对同一批次中的各项机械指标取平均值以减少误差对数据有效性的影响。2019 年中指标值区分了左右和果皮果肉，为了与 2018 年的数据对应，我们在有差异的指标上取了平均值，例如：2019 年的带皮硬度值等于（带皮硬度（左）+带皮硬度（右）/2）。由于各项指标的量纲不一，我们将数据进行标准化。处理后共得到 2018 和 2019 年的机械指标值数据共计 87 条，每一条对应 10 个机械指标值，数据文件见附件。下图为部分数据示例，上方为处理前原始值，下方为处理后数据。

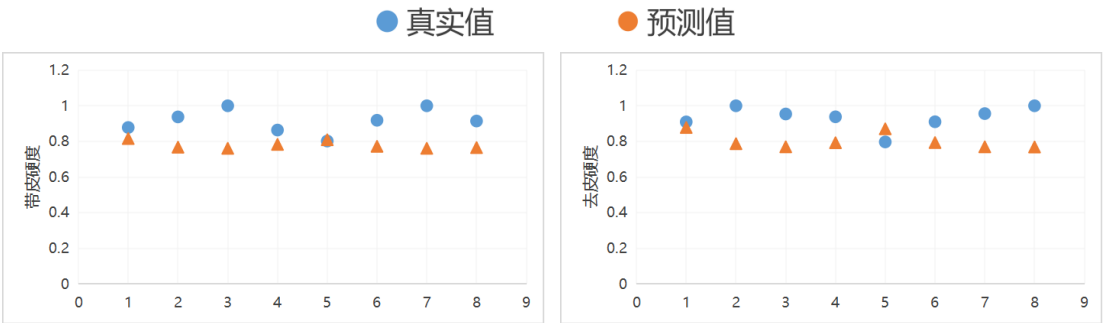
2018									
	带皮硬度	去皮硬度	TSS	L	a	b	c	h	
8 度	5.66375	4.83375	12.4625	83.8586864	-4.073965067	54.84562313	56.49651307	95.91603587	0
	4.75625	4.325625	12.44375	80.64225672	-2.809587523	57.39060159	57.46726706	92.80488805	3
	4.098125	3.449375	12.19375	79.3217582	-1.884603833	56.07641346	56.6242084	90.46686409	6
	4.019375	3.44125	11.93196389	81.68919368	-3.914479828	54.26618241	54.43264521	94.1459018	9
	2.39	1.694375	13.49375	79.4192008	-1.95790884	53.54322332	53.60023934	92.09758713	12
	带皮硬度	去皮硬度	TSS	L	a	b	c	h	
8 度	1	1	0.92357573	1	-1	0.955655135	0.983107706	1	0
	0.8397704	0.8948797	0.922186197	0.961644645	-0.689644481	1	1	0.967563841	0.25
	7	52							
	0.7235709	0.7136022	0.903659101	0.945897934	-0.462596955	0.977100987	0.985329759	0.943188105	0.5
	56	76							
8 度	0.7096667	0.7119213	0.884258556	0.974129183	-0.960852576	0.945558696	0.947193907	0.981544962	0.75
	4	86							
	0.4219819	0.3505301	1	0.94705992	-0.480590483	0.932961527	0.932709037	0.960189673	1
02		27							

步骤 1 中所获得的数据可看作时序数据，这是因为在相同年份、相同温度、相同批次的后序数据是由前序数据变化而得。因此我们利用 LSTM 模型可以学习出其中的变化规律。我们将每一批次中每一天的机械指标值对应为一个 10 维向量，下文中称作：指标值向量，这个 10 维向量项对应了 10 个指标值（带皮硬度，去皮硬度，...，h）。模型的输入为前一天的指标值向量，对应的标签值为后一天的指标值向量。

由于样本量较少，我们只选取了 15% 的样本作为测试集，模型采用基于 pytorch 架构的单向单层 LSTMCell，LSTMCell 中隐向量维度为 20 维，优化器为 SGD，学习率为 0.01，momentum 等于 0.9，损失函数为平方和损失函数。

3.4 模型预测结果分析

3.4.1 模型预测效果检验



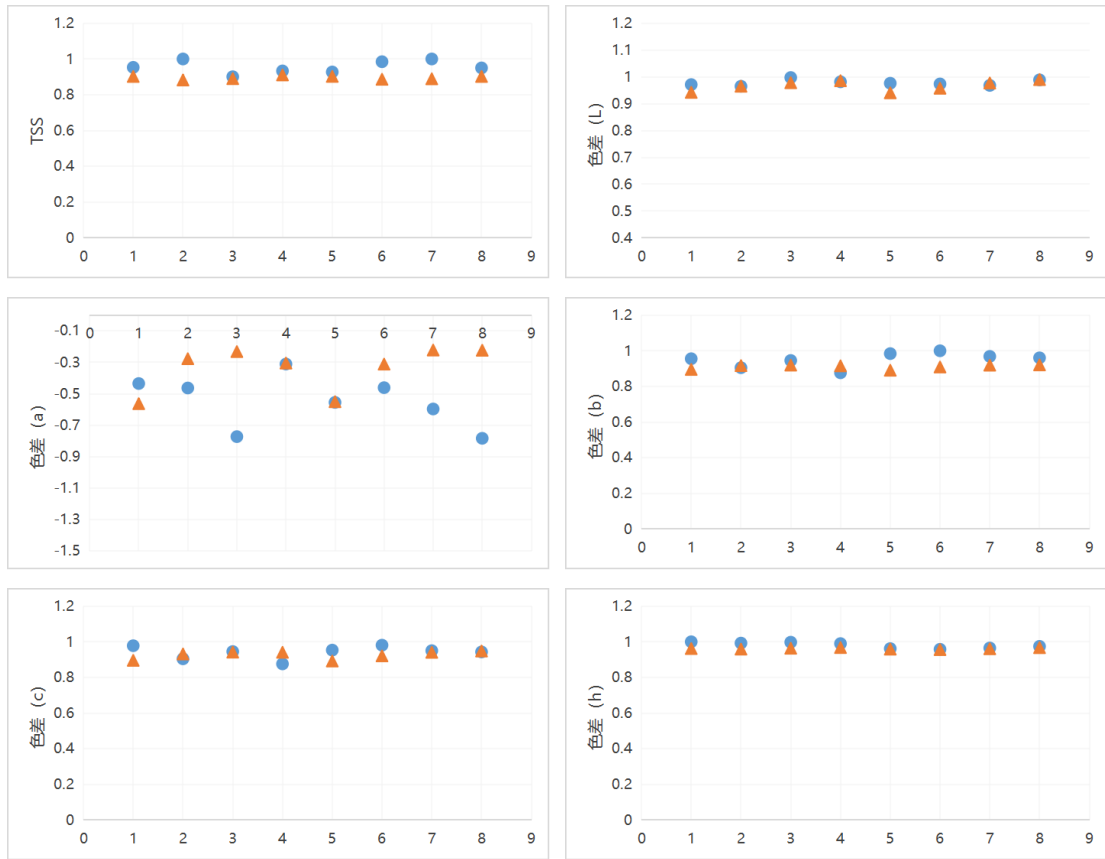


图 3.1 各项指标的模型预测值和真实值的对比

图 3.1 展示了测试集上各项指标的模型预测结果和真实结果的比对，从图上可以看出模型学习到的指标值变化规律与真实变化规律大体吻合，但是受限于样本数据量，预测模型并不能做到与真实值完全对应，在实际应用场景下，可以通过增加数据量来克服这一缺陷。

下方给出了模型训练的损失值变化情况。可以看出，在 26 个 Epoch 后，模型基本收敛，学习有效。

Epoch: 2, Loss: 0.1899154484

Epoch: 4, Loss: 0.1530237645

Epoch: 24, Loss: 0.0318585932

Epoch: 26, Loss: 0.0283844806

Epoch: 28, Loss: 0.0259200633

Epoch: 30, Loss: 0.0240719188

3.4.2 贮藏环境和贮藏时间对指标值的影响

接下来，利用训练好的模型来预测贮藏环境和贮藏时间对机械指标值的影响。我们从数据集中选取样本，通过改变贮藏环境和贮藏时间对应维度的数值来模拟 1 度、8 度及不同贮藏时间。例如：现有一批 1 度贮藏环境下的带皮硬度真实值，我们选取第 0 天的真实值，并将其贮藏环境维度的数值变为 8 度对应的数值，然后将改变后的指标值向量输入模型进行预测，后一天的输入为前一天的预测结果。通过这样的做法，可以模拟出真实条件下不能获取的数据，并能基于此分析贮藏环境和贮藏时间对指标值的影响。

下图中，蓝色线条表示真实的指标值（例如：带皮硬度）随时间的变化趋势，橙色线条为利用真实的第0天指标值向量预测得到的结果；灰色为使用改变贮藏环境维度的数值后指标值向量预测得到的结果。

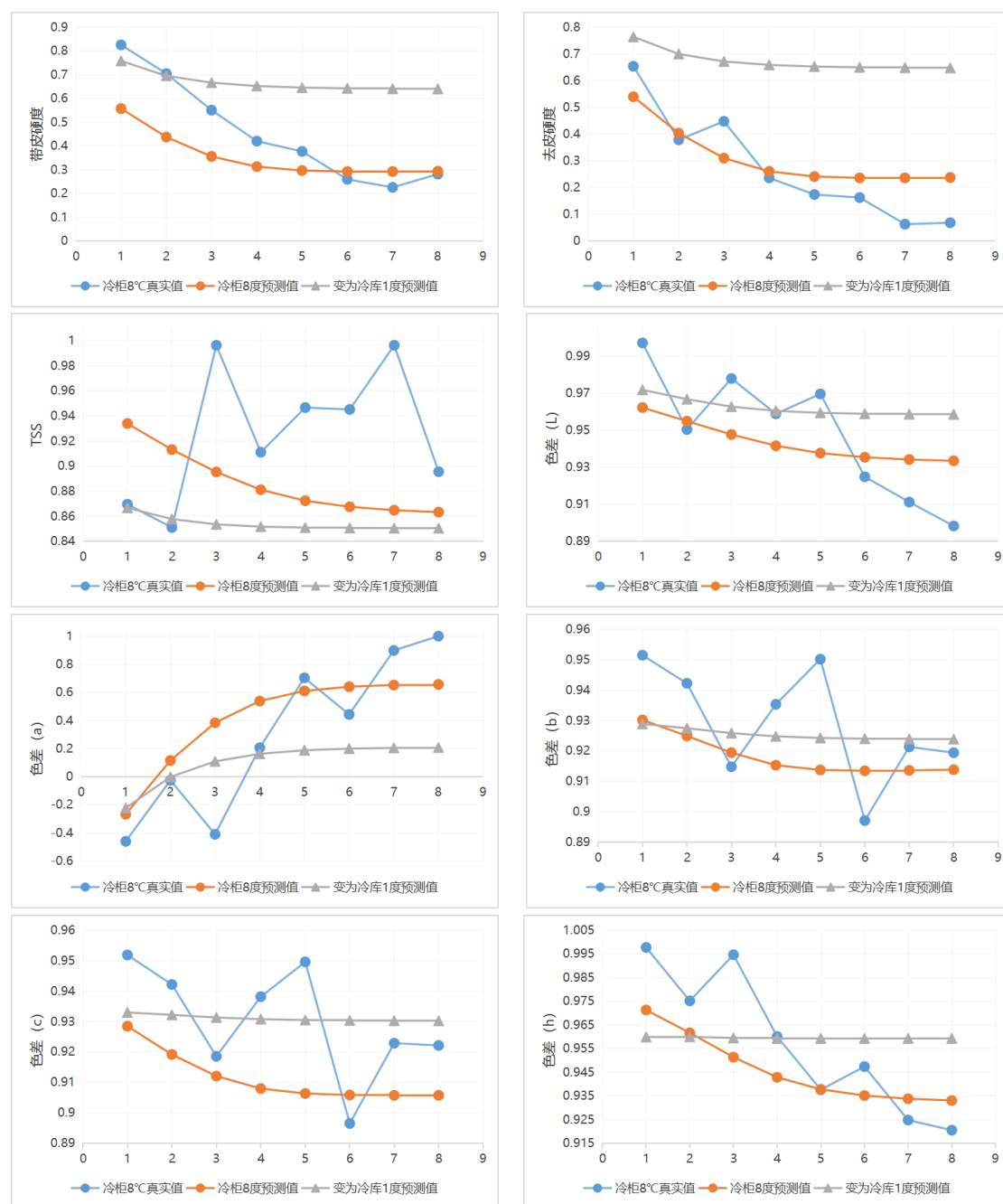


图 3.2 真实贮藏环境为 8°C 的预测结果

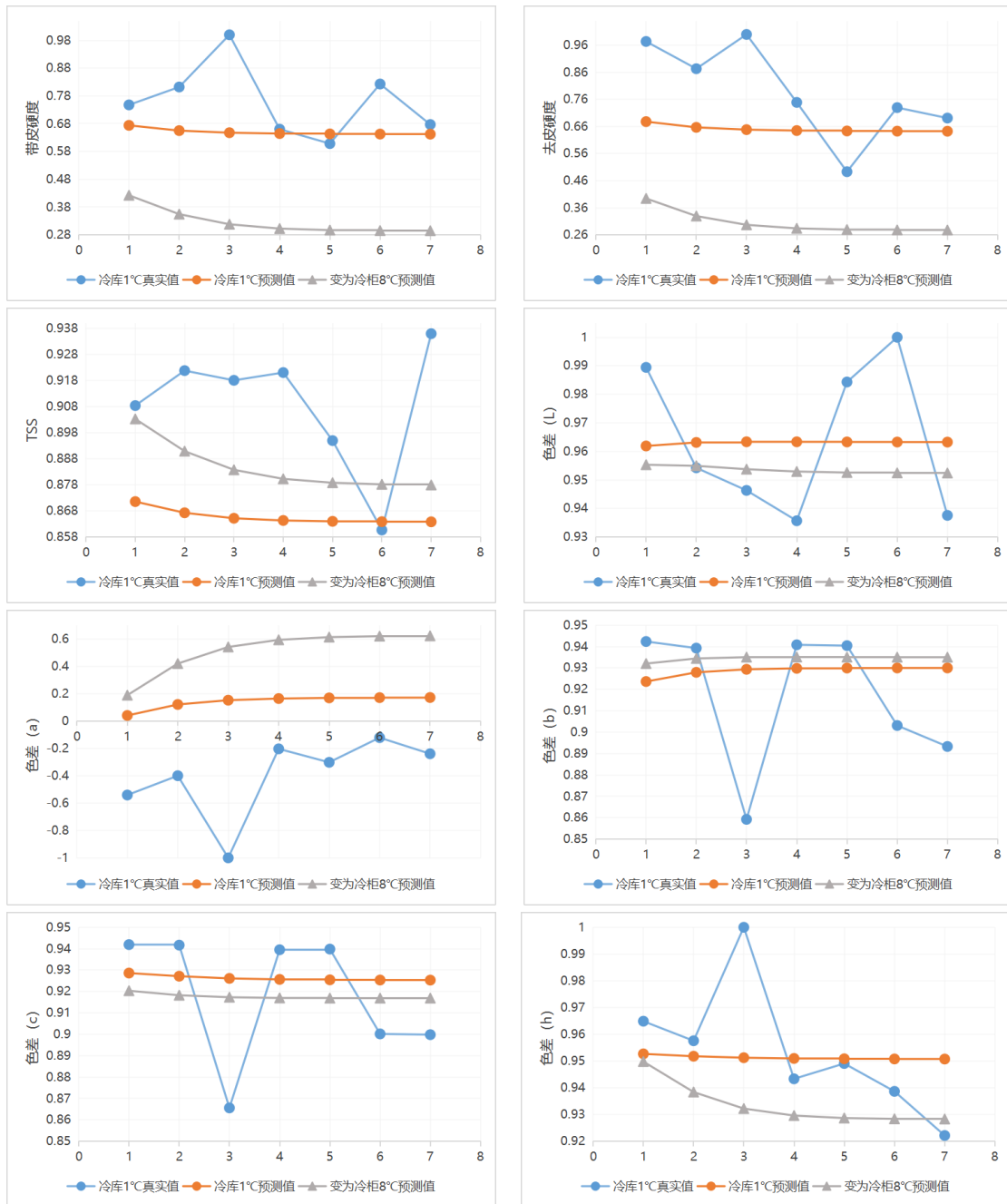


图 3.3 真实贮藏环境为 1°C 的预测结果

如图 3.2、图 3.3 所示，模型预测的指标值变化规律与真实结果大致吻合。结果表明，当贮藏环境从冷柜 8°C 变为冷库 1°C 时，带皮硬度和去皮硬度下降趋势明显变缓，而当贮藏环境从冷库 1°C 变为冷柜 8°C 时，带皮硬度和去皮硬度下降趋势明显变快。同时预测结果表明 TSS 指标值在冷库 1°C 变化平缓，而在冷柜 8°C 波动较大。TSS 指标值与黄桃人工评分中的味道有密切关系。综合模型的两个预测结果，可知在相对较低的贮藏环境下，黄桃能更长时间地保持较好的品质，与“普遍认为”一致。

3.4.3 下年度有损测试次数评估

本文模型主要基于往年数据学习而得，但是模型在预测当年指标值变化规律

时需要一个初始指标值作为基准（以此对后续待预测数据进行放缩），同时还需要少量当年样本数据对模型进行微调，以更好地进行预测。微调所需数据量与所需的效果有关，通常认为是训练模型所需数据量的 20%，对于本文模型而言也就是大约 $87 * 0.2 = 17$ 次。

4. 结语

为了更加科学地评判黄桃品质，需要综合利用这两种评判方法。本文通过线性回归、逐步线性回归方法，循环神经网络模型（LSTM），根据 2018 年至 2020 年的实验数据，分析了黄桃样本的各项机械指标评分与人工评分之间的关系以及两类方法的评判一致性，构建黄桃品质分析的定量模型。利用时序数据训练循环神经网络模型。该模型可以根据各项机械指标的初始值预测其变化趋势，从而选择最利于果农的贮藏方式和环境。

参考文献

- [1] 林震山, 陈观文, 郭吴珉, 王则金. 建宁黄桃最适贮藏温度研究[C]. 福建省制冷学会 2016 年学术年会, 2016.
- [2] 蒋均, 骆房增, 柳建良, 骆韵芳, 丘苑新, 汤梅, 王琴. 不同温度与包装对鹰嘴蜜桃贮藏效果的影响[J]. 现代食品科技, 2017, 33(10): 178-185+132.
- [3] 余意, 张鑫, 张丽平, 高中山, 贾惠娟. 采收成熟度和贮藏温度对锦绣黄桃完熟品质的影响[J]. 食品工业科技, 2015, 36(04): 334-338.
- [4] 吴剑, 褚伟雄, 王晖. 热处理-冷藏对锦绣黄桃生理生化的影响及感官评价[J]. 食品科技, 2015, 40(10): 314-319.
- [5] 游士兵, 严研. 逐步回归分析法及其应用[J]. 统计与决策, 2017(14): 31-35.
- [6] 许鑫. 不同品种绵羊体重与体尺指标的相关回归分析[D]. 甘肃农业大学, 2017.
- [7] 吴巧英. 薄丝织物缝纫平整度与织物结构及性能的相关性研究[D]. 东华大学, 2013.
- [8] 陆会宁, 刘妍妍, 张丽, 金鑫, 何林波, 曹忻. 兰州大尾羊×小尾寒羊 F1 代体重和体尺的相关性研究[J]. 中国畜牧兽医, 2019, 46(06): 1628-1635.
- [9] 李松臣, 张世英. 基于逐步回归法的人口出生率影响因素分析[J]. 统计与决策, 2008(04): 7-9.
- [10] 王红, 童恒庆, 魏平. 进出口贸易额预测的逐步回归建模研究[J]. 统计与决策, 2006(02): 80-82.
- [11] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural networks, 2015, 61: 85-117.
- [12] 王鑫, 吴际, 刘超, 杨海燕, 杜艳丽, 牛文生. 基于 LSTM 循环神经网络的故障时间序列预测[J]. 北京航空航天大学学报, 2018, 44(04): 772-784.
- [13] 彭燕, 刘宇红, 张荣芬. 基于 LSTM 的股票价格预测建模与分析[J]. 计算机工程与应用, 2019, 55(11): 209-212.
- [14] 王渊明. 基于 LSTM 神经网络的电商需求预测的研究[D]. 山东大学, 2018.

附录

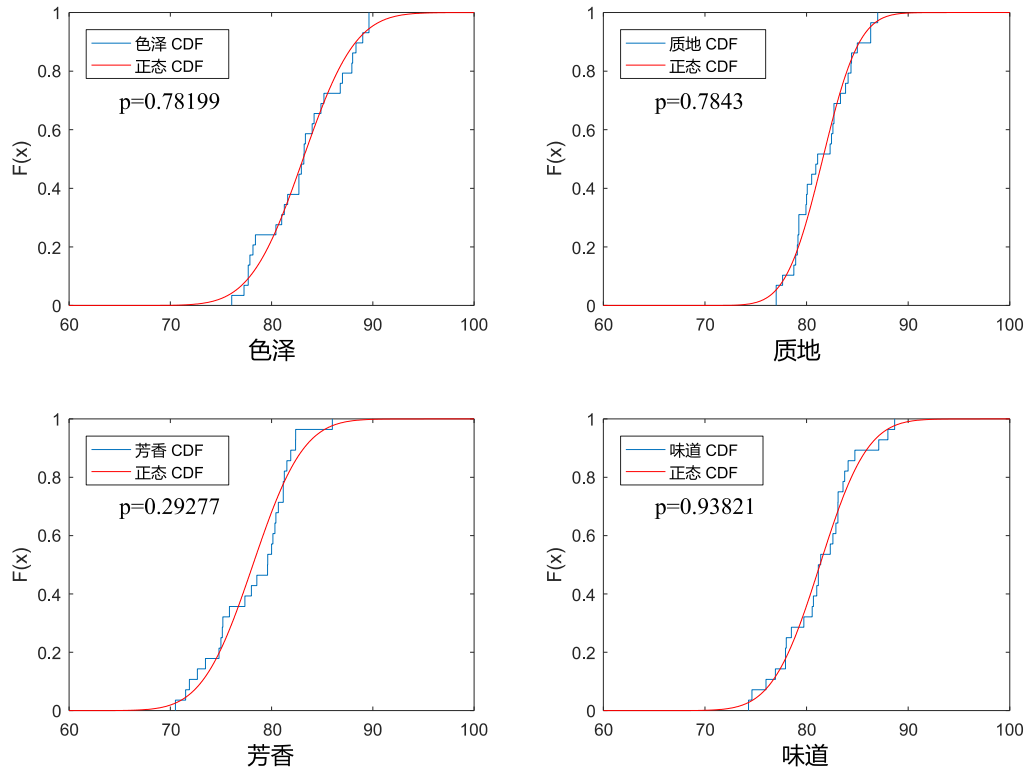


图 4.1 2018 年人工评分分布与正态分布的累积分布函数对比

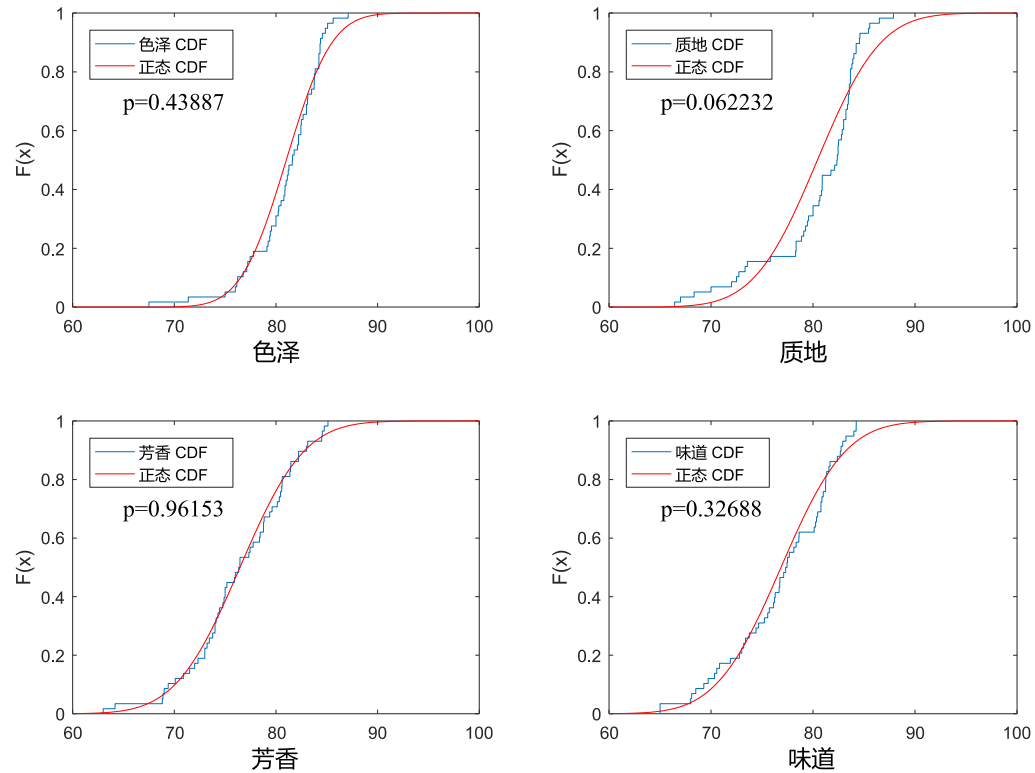


图 4.2 2019 年人工评分分布与正态分布的累积分布函数对比

程序

运行环境:

Matlab 2016a

Win10 操作系统

数据的正态性检验

```
clear
i=3; % file_3_mean.xls 2018 年的人工评分数据
%i=6, file_6_mean.xlsx 2019 年的人工评分数据
filename=['D:\modeling\file_',num2str(i),'_mean.xls'];
sheet = 'all';
en=31; % 2018 年数据量
% en=60, 2019 年数据量
% 读取人工评分的四类分数
man_color = xlsread(filename, sheet, ['B2:B',num2str(en)]); %色泽人工评分
man_texture = xlsread(filename, sheet, ['C2:C',num2str(en)]); %质地人工评分
man_smell = xlsread(filename, sheet, ['D2:D',num2str(en)]); %芳香人工评分
man_taste = xlsread(filename, sheet, ['E2:E',num2str(en)]); %味道人工评分

for i=1:1:1 %去除一组异常数据
    man_color(man_color==min(man_color))=[];
    man_texture(man_texture==min(man_texture))=[];
    man_smell(man_smell==min(man_smell))=[];
    man_taste(man_taste==min(man_taste))=[];
end

% K-S 检验色泽人工评分数据是否正态
[mu1, sigma1] = normfit(man_color);
test_cdf = makedist('Normal','mu',mu1,'sigma',sigma1);
[h, p1] = kstest(man_color,'CDF',test_cdf,'Alpha',0.05);

% K-S 检验质地人工评分数据是否正态
[mu2, sigma2] = normfit(man_texture);
test_cdf = makedist('Normal','mu',mu2,'sigma',sigma2);
[h, p2] = kstest(man_texture,'CDF',test_cdf,'Alpha',0.05);

% K-S 检验芳香人工评分数据是否正态
[mu3, sigma3] = normfit(man_smell);
test_cdf = makedist('Normal','mu',mu3,'sigma',sigma3);
[h,p3] = kstest(man_smell,'CDF',test_cdf,'Alpha',0.05);

% K-S 检验味道人工评分数据是否正态
[mu4, sigma4] = normfit(man_taste);
```

```

test_cdf = makedist('Normal','mu',mu4,'sigma',sigma4);
[h,p4] = kstest(man_taste,'CDF',test_cdf,'Alpha',0.05);
% 人工评分四类评分数据的 CDF 与正态 CDF 可视化
figure(1)
subplot(2,2,1)
cdfplot(man_color)
hold on
x_values = 60:0.001:100;
y = normcdf(x_values, mu1,sigma1);
plot(x_values,y,'r-')
s=['p=',num2str(p1)];
text(65, 0.7, s,'FontSize', 12, 'Fontname', 'Times New Roman' )
xlabel('Color'); % 色泽人工评分
legend('Color CDF','Normal CDF','Location','NorthWest','Box','off')

% 人工评分数据的 CDF 与正态 CDF 可视化
subplot(2,2,2)
cdfplot(man_texture)
hold on
y = normcdf(x_values, mu2,sigma2);
plot(x_values,y,'r-')
s=['p=',num2str(p2)];
text(65, 0.7, s,'FontSize', 12, 'Fontname', 'Times New Roman' )
xlabel('Texture');
legend('Texture CDF','Normal CDF','Location','NorthWest','Box','off')

subplot(2,2,3)
cdfplot(man_smell)
hold on
% y = gamcdf(x_values, mu3,sigma3);
y = normcdf(x_values, mu3,sigma3);
plot(x_values,y,'r-')
s=['p=',num2str(p3)];
text(65, 0.7, s,'FontSize', 12, 'Fontname', 'Times New Roman' )
xlabel('Smell');
legend('Smell CDF','Normal CDF','Location','NorthWest','Box','off')

subplot(2,2,4)
cdfplot(man_taste)
hold on
% y = gamcdf(x_values, mu4,sigma4);
y = normcdf(x_values, mu4,sigma4);

```

```

plot(x_values,y,'r-')
s=['p=',num2str(p4)];
text(65, 0.7, s,'FontSize', 12, 'Fontname', 'Times New Roman' )
xlabel('Taste');
legend('Taste CDF','Normal CDF','Location','NorthWest','Box','off')

pdf_name = 'cdf_2018_file_3_1';
set(gcf, 'position',[10,10,920,650]);
saveas(gcf,['D:\modeling\',pdf_name,'.pdf']);

man_color = round(man_color);
man_texture = round(man_texture);
man_smell = round(man_smell);
man_taste = round(man_taste);
x_color = min(man_color):1:max(man_color);
for num=1:length(x_color)
    color(num)=length(find(man_color==x_color(num)));
end
x_texture = min(man_texture):1: max(man_texture);
for num=1:length(x_texture)
    texture(num)=length(find(man_texture==x_texture(num)));
end
x_smell = min(man_smell):1: max(man_smell);
for num=1:length(x_smell)
    smell(num)=length(find(man_smell==x_smell(num)));
end
x_taste = min(man_taste):1: max(man_taste);
for num=1:length(x_taste)
    taste(num)=length(find(man_taste==x_taste(num)));
end

figure(2)
subplot(2,2,1)
bar(x_color,
color,'FaceColor',[40/255,102/255,0.8],'EdgeColor',[55/255,102/255,0.8])
xlabel('Color');
axis([min(x_color)-2,max(x_color)+2,0,max(color)+1])
hold on;

subplot(2,2,2)
bar(x_texture,texture,'FaceColor',[40/255,102/255,0.8],'EdgeColor',[55/255,102/25
5,0.8])
xlabel('Texture');

```

```

axis([min(x_texture)-2,max(x_texture)+2,0,max(texture)+1])
hold on;

subplot(2,2,3)
bar(x_smell,smell,'FaceColor',[40/255,102/255,0.8],'EdgeColor',[55/255,102/255,0.8])
xlabel('Smell');
axis([min(x_smell)-2,max(x_smell)+2,0,max(smell)+1])
hold on;

subplot(2,2,4)
bar(x_taste,taste,'FaceColor',[40/255,102/255,0.8],'EdgeColor',[55/255,102/255,0.8])
xlabel('Taste');
axis([min(x_taste)-2,max(x_taste)+2,0,max(taste)+1])
hold on;

pdf_name = 'distr_2018_file_3_1';
set(gcf, 'position',[10,10,920,650]);
saveas(gcf,['D:\modeling\',pdf_name,'.pdf']);

```