

2026 年基于 Python 语言和 R 语言的数据 技术系列培训和应用技术服务预告

数智化时代，数据作为第五大生产要素，正成为驱动高质量发展的引擎。“工欲善其事，必先利其器”。为助力相关专业技术人员提升数据技能，赋能相关机构更好挖掘和深度开发数据价值，广东省公共卫生研究院依托Python语言和R语言双平台，提供从能力建设到实践应用的全周期数据技术服务，现推出 2026 年数据技术系列培训和应用技术服务，具体预告如下：

一、基于Python语言和R语言的数据技术系列培训

2026 年，我院计划开展基于Python语言和R语言的数据技术培训班共有 6 期，其中Python语言与R语言各 3 期（课程安排见附件 1、2）。该系列培训内容设计主要聚焦医学数据全流程处理技术，并按照基础班、高级班和专题班的课程设计理念分期，以满足不同层次的数据技术需求，以帮助学员逐步掌握这两种主流编程语言在**医学数据治理、生物统计、智能预测、数据可视化及机器学习等前沿领域的应用范式**。教学方面，主要采用理论讲授与案例实操相结合的方式，各期培训班将主要围绕卫生健康领域的实践案例展开，使学员能够将所学方法和技能直接应用于该领域的数据分析和数据挖掘。

Python语言数据技术培训班将为学员提供正式出版的自编教材 2 册（详见附件 3、附件 4），以及教学案例数据和

程序文件；R语言数据技术培训班将为学员提供专用讲义，以及实践数据和代码。

本系列培训班接收预报名，有意愿参加学习者，请您扫描下方二维码填写预报名回执。同时，请您关注我们后续发布的各期培训班正式通知。届时，我们会提前将培训班的正式通知通过电子邮件发送给您，请注意查收。您也可关注广东省公共卫生研究院官方网站（<http://www.gdiph.org.cn>）“通知公告”栏目，或者通过“华南预防医学”官方网站（<http://www.hnyfyx.cn>）“编辑部公告”栏目和其微信公众号查询相关资讯；您也可致电 020-31051612，咨询沈老师或容老师，咨询 QQ 群：801116971。

欢迎各单位及相关人员踊跃报名、积极参加。



预报名回执二维码

二、基于Python语言和R语言的数据技术应用实践服务

除以上系列培训外，我院还致力于为个人和专业机构提供有针对性的数据技术支持和专题解决方案，促进智能数据技术在个人职业发展与机构业务价值提升中的转化应用。一是接收相关专业技术人员数据技术专题学习进修；二是承接数据技术咨询服 务，提供解决方案；三是承接专业机构数据技术能力建设指导、数据治理和分析等业务委托服务项目。

相关事宜可随时与我们联系，致电 020-31051632 咨询肖老师，或致电 020-31051979 咨询曾老师。

欢迎各单位及相关人员垂询。

附件：

1. 2026 年基于Python语言的数据技术系列培训计划表
2. 2026 年基于R语言的数据技术系列培训计划表
3. 《Python卫生健康统计分析与可视化——方法与实践》
书讯
4. 《Python卫生健康机器学习基本方法与实践》书讯

广东省公共卫生研究院
《华南预防医学》编辑部

2026 年 2 月 6 日

附件 1

2026 年基于 Python 语言的数据技术系列培训计划表

举办时间	培训班名称	课程单元	课程专题及主要内容	举办地点
5 月	基层业务人员适宜数据技能提升培训班	1	公共卫生数据的科研价值挖掘技术与实践 (1) 公共卫生数据来源与种类 (2) 数据分析工具 (3) 数据分析模式 (4) 现代数据分析方法两大类：统计学与机器学习 (5) 数据分析方法与科研实践案例 ①基于 Joinpoint Regression 模型和 JRP 软件的疾病监测数据统计分析研究 ②基于 Python 语言的公共卫生数据统计分析与机器学习研究 (6) 关于公共卫生科研的几点思考	佛山
		2	Python 语言基础及其数据分析技术概述 (1) Python 软件安装及运行 (2) 数据类型结构 (3) 基础语法介绍 (4) 数据导入导出 (5) 统计分析相关库概述	
		3	Python 数据可视化方法与医学实践 (1) Python 绘图方式及基本步骤 (2) Matplotlib 绘图库概述与实践，包括折线图、散点图、饼图、箱式图、玫瑰图等 (3) Seaborn 绘图库概述与实践，包括热力图、聚类热力图等	
		4	Python 数据治理方法与实践 (1) Numpy 库功能、方法及其应用实践（数组创建、数组对象索引、函数及其应用） (2) Pandas 库功能、方法及其应用实践（Series 创建、DataFrame 创建、索引、修改和过滤） (3) 数据清理（统计描述、缺失值和异常值处理）和数据转化（生成新变量、变量分组）	
		5	Python 统计分析方法与实践 (1) Python 统计分析步骤及技术路线 (2) 假设检验：t 检验（两独立样本体重差异、配对样本血压差异检验）、卡方检验（分性别的肥胖率差异、高血压视网膜病变检出率差异、视网膜中央及分支静脉阻塞检出率差异检验） (3) 相关分析：Pearson、Spearman 相关系数及相关矩阵（以身高和体重的相关分析为例） (4) 回归分析：线性回归、Logistic 回归（以健康体检数据进行体重回归和体质类别分析为例）	

		6	Joinpoint 回归模型和 JRP 软件数据分析与科研实践 (1) JPR 模型的基本原理和分析方法 (2) JPR 模型的实现程序: Joinpoint Regression Program (JRP) 软件功能和使用方法 (3) 基于 JPR 模型和 JRP 软件的科研实践案例 ①死因监测数据分析 ②成人肺炎疾病负担分析 ③传染病发病率趋势分析 ④疫苗接种效果评价 AI 辅助 Python 数据分析	
8 月	卫生健康适宜机器学习方法与实践能力建设培训班	1	Python 语言基础概述和可视化基础与实践 (1) Python 软件安装及运行 (2) 数据类型结构 (3) 基础语法介绍 (4) 数据导入导出 (5) Python 绘图基础 (6) Matplotlib 库绘图技术介绍与实践, 包括线图、散点图和三维散点图。	广州
		2	机器学习概述与模型评估 (1) 机器学习在卫生健康领域的应用 (2) 机器学习 Scikit-learn 库概述 (3) 机器学习模型评估和模型选择方法	
		3	机器学习分类方法实践与结果可视化 (1) Logistic 回归二分类和多分类 (分别以肥胖和体质四分类预测为例) (2) 朴素贝叶斯分类 (以体质分类预测为例) (3) 随机森林分类 (以高血压预测为例) (4) 支持向量机分类 (以肥胖预测为例)	
		4	机器学习回归方法实践与结果可视化 (以体质指数预测为例) (1) 普通线性回归 (2) 多项式回归 (3) 岭回归 (4) 最近邻回归	
		5	机器学习聚类方法、降维方法实践与结果可视化 (1) K-Means 聚类 (以身高和体重为特征变量进行聚类分析) (2) 层次聚类 (以身高、体重和腰围为特征变量进行聚类分析) (3) 主成分分析 (以成年人的体质测量值主成分分析为例) (4) 线性判别降维和随机森林分类 (以体质测量指标降维后进行体质分类为例)	
		6	机器学习图像处理与图像识别 (1) 图像处理基本知识及相关库用法 (2) 图像数字化处理 (3) 图像分类 ①采用单一分类算法进行图像分类, 以 5 种食材的多层感知器分类为例	

			<p>②从多种学习算法中选择最佳模型进行图像分类，采用 5 种分类模型的工作流机器学习对 5 种食材进行识别</p> <p>AI 辅助 Python 机器学习</p>	
11 月	公共卫生数据挖掘利用高级技术研讨班	1	<p>Python 语言及统计分析和可视化技术基础</p> <p>(1) Python 软件安装及运行</p> <p>(2) 数据类型结构</p> <p>(3) 基础语法介绍</p> <p>(4) 数据导入导出</p> <p>(5) 统计分析相关库概述</p> <p>(6) Python 绘图基础</p> <p>(7) Matplotlib 库绘图技术介绍与可视化实践</p>	中山
		2	<p>回归分析和生存分析与结果可视化</p> <p>(1) 线性回归 (以健康体检数据进行体重回归分析为例)</p> <p>(2) Logistic 回归 (以健康体检数据进行体质类别分析为例)</p> <p>(3) 生存分析 (以两种治疗方案治疗肿瘤患者的生存时间分析为例)</p> <p>(4) Cox 比例风险回归 (以肿瘤患者生存时间的影响因素分析为例)</p>	
		3	<p>判别分析和聚类分析与结果可视化</p> <p>(1) Fisher 判别和 Bayes 判别 (以体质分类、疾病判别分析为例)</p> <p>(2) 层次聚类和 K-means 聚类 (以体检数据聚类分析为例)</p>	
		4	<p>主成分分析和主元回归分析与结果可视化</p> <p>(1) 主成分分析 (以复方药物化合物的主成分分析为例)</p> <p>(2) 主元回归分析 (以体检数据为例进行体质指数的主成分回归分析)</p>	
		5	<p>时间序列趋势分析预测与结果可视化</p> <p>(1) ARIMA 模型 (以手足口病月发病数变化趋势分析预测为例)</p> <p>(2) SARIMA 模型 (以考虑季节性的手足口病发病数趋势分析预测为例)</p> <p>(3) SARIMAX 模型 (以考虑季节性和协变量的手足口病发病数趋势分析预测为例)</p> <p>(4) VARMAX 模型 (以考虑协变量的手足口病发病数和发病率趋势向量模型分析预测为例)</p>	
		6	<p>传染病动力学模型分析与结果可视化 (以传染病流行特征及趋势分析预测为例)</p> <p>(1) 基本再生数和实时再生数计算</p> <p>(2) SEIR 模型分析</p> <p>(3) SEEIR 模型分析</p> <p>(4) SSEIR 模型分析</p> <p>AI 辅助 Python 编程</p>	

附件 2

2026 年基于 R 语言的数据技术系列培训计划表

举办时间	培训班名称	课程单元	课程专题及主要内容	举办地点
4 月	医学健康实用统计分析和可视化技术培训班	1	R 语言基础及 AI 辅助编程 (1) R 语言基础: R 软件及 R 程序包安装、数据结构、编程基础、数据导入与导出等 (2) AI 工具辅助技术解答、代码解读和代码生成等	佛山
		2	数据管理技术及实践 数据管理基础: 数据合并、数据匹配、变量重编码、缺失值基本处理等, 以居民健康档案数据清理和质量控制为例	
		3	绘图基础与 ggplot2 绘图 (1) 绘图基础, ggplot2 绘图体系和基础语法, 涵盖几何对象、统计变换等核心函数, 图形美化 (图例定制、坐标轴设置、配色方案、主题设置等) (2) 常用图形可视化全流程制图技巧, 包括线图、条图、散点图、箱式图、误差棒图、热力图、分面图、交互式图形等的绘图技巧和优化	
		4	常用统计分析方法 (1) 描述性统计: 均值、中位数、构成比等指标的便携批量分析 (2) 常用统计分析: 正态性检验; t 检验、卡方检验、方差分析; 相关分析等方法的分析和实践示例 (健康调查、临床体检等数据为例) (3) 综合实践: 以健康调查体检数据为例, 完成描述性统计、组间比较等	
		5	相关分析和回归分析 (1) 相关分析 (以健康体检数据的相关分析为例) (2) 线性回归 (以健康体检数据进行体重回归分析为例) (3) Logistic 回归分析 (以狂犬病暴露者被动免疫制剂使用的影响因素分析为例)	
		6	经典时间序列模型分析 自回归差分移动平均模型 (ARIMA)、SARIMA、SARIMAX 的原理、建模步骤、注意事项与实践 (以医院门诊量预测, 基于外生变量 (影响因素) 的传染病流行趋势预测为例)	
		1	R 语言基础及 AI 辅助编程 (1) R 语言基础: R 软件安装、程序包安装、数据结构、编程基础、数据导入与导出等 (2) AI 工具辅助技术解答、代码解读和代码生成等	

7 月	健康数据分析利用高级培训班	2	数据管理技术 (1) 数据类型转换、数据重构、变量重编码、条件与循环语句、缺失值处理等数据管理技巧及描述性统计分析 (2) 案例综合实践：以卫生健康相关数据的前处理及转换为时间序列数据为例	广州
		3	ggplot2 绘图及时空数据可视化 (1) 绘图基础语法、ggplot2 绘图系统介绍 (2) 基于疾病监测数据的线图、条图、散点图；时序图；疾病发病率空间分布地图、六边形地图（邮票地图）绘制	
		4	回归分析和生存分析 (1) Logistic 回归分析（以乳腺癌复发的影响因素分析为例） (2) 生存分析（以两种治疗方案治疗肿瘤患者的生存时间分析为例） (3) Cox 比例风险回归（以肿瘤患者生存时间的影响因素分析为例）	
		5	年龄 - 时期 - 队列（APC）效应分析及疾病趋势预测 (1) 年龄时期队列（APC）分析模型的基本原理和分析步骤 (2) APC 效应分析：基于基本负担和疾病报告数据，分析某疾病的年龄、时期、队列效应 (3) 贝叶斯年龄-时期 - 队列（BAPC）分析模型原理和分析步骤 (4) BAPC 疾病流行趋势预测：构建 BAPC 模型，预测某慢性疾病未来 20 年不同人群的发病趋势	
		6	贝叶斯结构时间序列模型的建模及应用 (1) 贝叶斯结构时间序列模型（BSTSM）的基本原理和建模步骤 (2) BSTSM 在传染病流行趋势预测中的应用（以登革热流行趋势分析为例） (3) BSTSM 疫苗接种策略效果评价中的应用（以非药物干预措施效果评估为例）	
10 月	机器学习疾病风险预测与评估方法培训班	1	R 语言概述及 AI 辅助编程 (1) R 语言基础：R 语言简介，R 包安装、数据结构、编程基础、数据导入导出等 (2) AI 工具辅助 R 语言技术解答、代码解读、代码生成等	珠海
		2	数据管理及可视化技术 (1) 数据管理技术及实践（基于 tidyverse 包的疾病监测数据管理和批量处理等） (2) ggplot2 绘图语法和绘图技巧，绘制条图、折线图、散点矩阵图、箱式图、热力图、时序图、森林图等	
		3	机器学习概述与模型评估 (1) 机器学习定义、核心特性与分类，优势和价值，以及其卫生健康领域的应用场景 (2) 机器学习常用方法和基本分析流程 (3) 机器学习模型评估和选择，核心评估指标，模型选择方法，机器学习全流程工作流等	

		4	<p>机器学习之聚类与降维</p> <p>(1) 聚类分析: K-means 聚类和层次聚类原理和实践 (案例: 基于某区域社会经济卫生综合指标的聚类分析)</p> <p>(2) 主成分分析原理、实践和结果解读 (案例: 药物化合物成分质量评价指标的主成分分析)</p> <p>(3) 因子分析原理、实践和结果解读 (案例: 健康体检指标体系的因子分析与维度提取)</p>	
		5	<p>机器学习之分类预测模型</p> <p>(1) 决策树分类原理、模型构建、模型评估、预测应用 (案例: 乳腺癌复发风险的预测分析)</p> <p>(2) 支持向量机分类原理、模型构建、模型评估、预测应用 (案例: 儿童百日咳重症的影响因素及预测分析)</p> <p>(3) SHAP 可解释性分析原理, 分析上述支持向量机模型研究中特征变量的贡献, 绘制蜂群图和瀑布图</p>	
		6	<p>机器学习之回归预测模型</p> <p>(1) 多项式回归预测 (案例: 年龄、BMI 等与血压/血糖的非线性关系与预测分析)</p> <p>(2) 最近邻回归 (KNN) 预测 (案例: 基于健康体检资料的血压/血糖值预测分析)</p> <p>(3) 随机森林回归预测 (案例: 登革热传播的环境影响因素重要性排序及其流行趋势预测分析)</p>	

附件 3

《Python 卫生健康统计分析与可视化——方法与实践》书讯

《Python 卫生健康统计分析与可视化——方法与实践》由广东省公共卫生研究院曾四清主任医师著，于 2024 年 12 月由中山大学出版社正式出版发行。ISBN:978-7-306-08180-3。该书主要介绍了 Python 语言在卫生健康领域数据治理、统计分析、预测和可视化等方面的方法与实践，内容分为六章，共计 77.9 万字。第一章介绍了 Python 语言基础；第二章介绍了 Python 统计分析相关库；第三章介绍了 Python 数据可视化库及方法实践，含 70 多种图形绘制方法的 Python 语言实践，既包括结果展示类图形，也包括统计分析类图形及复合图形。第四章、第五章和第六章，从基本的数据清理、描述性分析到高级统计与预测模型分析，全面系统介绍了 100 多种统计分析和预测模型的 Python 语言实践。

该书作为一本数据处理“工具书”，通俗易懂，易学好用，其中实践案例具有较好的专业代表性、实用性和程序可复用性，可作为进行 Python 语言数据治理、统计分析、预测和可视化等技术学习、培训的教材，也可以作为相关技术应用实践的参考书。

该书由广东省公共卫生研究院资助出版。相关技术和业务咨询可联系广东省公共卫生研究院的曾老师（电话：020-31051979）。

附件 4

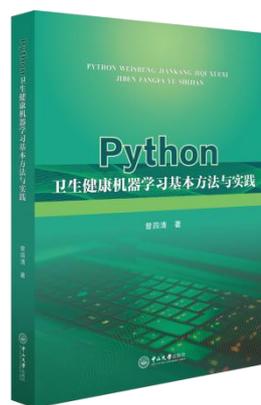
《Python 卫生健康机器学习基本方法与实践》书讯

《Python 卫生健康机器学习基本方法与实践》由广东省公共卫生研究院曾四清主任医师著，于 2025 年 5 月由中山大学出版社正式出版发行。ISBN: 978-7-306-08373-9。该书主要介绍了 Python 语言机器学习模型分析、评估和选择基本方法，以及 30 多种机器学习基本模型算法在卫生健康领域的实践，内容分为八章，共计 38.6 万字。第一章概要介绍了机器学习 Scikit-learn 库；第二章主要介绍数据预处理；第三章主要介绍模型评估和模型选择方法；第四章、第五章、第六章和第七章分别主要介绍分类、回归、聚类和降维等 4 大类机器学习模型应用实践；第八章主要介绍图像处理和图像识别技术。其中，分类算法包括逻辑回归分类、线性判别分类、朴素贝叶斯分类、k 近邻分类、决策树分类、随机森林分类、支持向量机分类和多层感知机分类 8 种；回归算法包括线性回归、多项式回归、岭回归、Lasso 回归、K 近邻回归、决策树回归、随机森林回归、支持向量回归和多层感知机回归 9 种；聚类算法包括 K-Means 聚类、加速 K-Means 聚类、谱聚类、均值转换聚类、DBSCAN 聚类、层次聚类和近邻传播算法 7 种；降维算法包括主成分分析、核主成分分析、独立成分分析、非负矩阵分解、截断奇异值分解、线性判别分析、因子分析和特征选择等 8 种。本书的案例实践中，还应用了多种 Python 语言可视化技术。

该书通俗易懂，各案例的实践程序均具有较好的可复用性，易学好用。主要读者对象包括卫生健康行业从业人员、高等院校预防医学和医学等专业的本科生和研究生，以及其他对机器学习预测和评估技

术有需求的人员。该书可以作为 Python 语言机器学习等技术学习、培训的入门级教材，也可以作为相关技术应用实践的参考书。

该书由广东省公共卫生研究院资助出版。相关技术和业务咨询可联系广东省公共卫生研究院曾老师（电话：020-31051979）。



两部 Python 语言数据分析技术的专著