

2025 年基于 R 语言和 Python 语言的数据 技术系列培训和应用技术服务预告

当今社会已经进入数智化时代，数据已成为第五大生产要素。“工欲善其事，必先利其器”。数据技术已成为推动各领域高质量发展的中坚力量。为助力提升相关工作者的数据技能和为相关机构的数据价值开发赋能，我院精心策划了 2025 年基于 R 语言和 Python 语言的数据技术系列培训服务，并提供相关数据技术能力建设和应用实践服务。

本系列培训包含 4 期 R 语言数据技术培训班（详见附件 1）和 3 期 Python 语言数据技术培训班（详见附件 2），旨在通过系统的课程设置和实践案例讲解，帮助学员掌握这两种主流编程语言在**数据清理、数据分析、预测、可视化和机器学习**等方面的强大应用功能与实践技能。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。

R 语言数据技术培训班将为学员提供专用讲义，Python 语言数据技术培训班则为学员提供正式出版的自编教材 2 册（详见附件 3、附件 4）。

同时，我院还接收相关数据技术专题学习进修，并开展专业机构数据技术能力建设和应用实践服务，致力于为有相应需求者提供有针对性的数据技术支持和专题解决方案，让智能数

据技术的应用实践更好为个人和专业机构数据技术能力提升和业务的高质量发展赋能。

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，联系邮箱：1991125464@qq.com，咨询 QQ 群：801116971。

欢迎大家咨询。



预报名回执二维码

附件：

1. 2025 年基于R语言的数据技术系列培训计划表
2. 2025 年基于Python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
2025 年 2 月 28 日

附件 1

2025 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
3 月 26-28 日	R 语言医学统计分析技术培训班	1	R 语言入门与基础知识 （R 软件安装、程序包安装、数据结构、创建数据集、基础语法） AI 工具辅助 R 语言代码编写和代码解读示例	广州
		2	数据管理及应用 （变量重编码、数据类型转换、数据合并、数据重构等，以卫生健康相关数据的处理为例）	
		3	绘图基础和 ggplot2 绘图 （绘图基础语法，ggplot2 绘图系统，线图、条图、箱式图、误差棒图、热力图、词云图、分面图等制图技巧）	
		4	统计分析（1） （描述性分析、t 检验、卡方检验、方差分析、相关分析等，以健康调查、临床体检等数据的统计分析为例）	
		5	统计分析（2） （多重线性回归、Logistic 回归，以疾病发病/预后的影响因素分析为例）	
		6	经典时间序列分析 （ARIMA、SARIMA、SARIMAX 建模与实践，以医院门诊量预测和手足口病流行趋势预测为例）	
6 月	R 语言医学数据挖掘技术培训班	1	R 语言数据处理 （数据导入与初步探索、数据合并与重塑、数据操作，基于卫生健康数据的高效数据清理技巧） AI 工具辅助 R 语言代码编写和代码解读示例	东莞
		2	ggplot2 绘图及交互式可视化 （绘图基础语法、ggplot2 绘图系统，基于疾病监测数据的时序图、热力图、误差图等制图技巧）	
		3	回归分析方法： 广义线性模型（以臭氧暴露与人群死亡关系分析为例） 广义相加模型（以登革热发病趋势预测为例） 混合线性模型（以小学生肺功能的影响因素分析为例）	
		4	机器学习分类方法 （以乳腺病变分类分析为例）： Logistic 回归 支持向量机 随机森林	
		5	机器学习聚类和降维方法： K-means 聚类（以某省社会卫生经济学指标聚类分析为例） 主成分分析（以某种药物化合物的成分质量评价指标分析为例） 因子分析（以健康体检指标数据因子分析为例）	

		6	Meta 分析 （分类型变量 Meta 分析、连续型变量 Meta 分析，以多地区/多中心的癌症发生风险 RR 值、 β 值等综合效应评估为例，以多项研究的 EV71 疫苗接种综合效果评估为例）	
9 月	基于 R 语言的疾病预测与风险评估技术培训班	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） AI 工具辅助 R 语言代码编写和代码解读示例	广州
		2	数据管理技术及实践 （基于 tidyverse 包的疾病监测、个案调查数据管理、批量操作等）	
		3	R 语言可视化技术 （绘图基础，ggplot2 绘制线图、条图、箱式图、误差棒图、热力图，空间数据绘图等）	
		4	时间序列模型： ARIMA 模型（以手足口病流行趋势分析预测为例） SARIMAX 模型（以登革热流行趋势分析预测为例）	
		5	生存分析和回归分析： Logistic 回归分析（以维生素 D 水平与骨质疏松症的关系分析为例） 生存分析（以不同治疗方案治疗肿瘤患者的生存时间分析为例） Cox 比例风险回归分析（以肺癌患者生存时间的影响因素分析为例）	
		6	基于机器学习的疾病风险评估： 随机森林（以登革热影响因素重要性及流行趋势预测分析为例） 支持向量回归（以脊椎病变患者的发病风险评估为例）	
10 月	基于 R 语言的居民健康档案数据质控及分析技术培训班	1	专题报告： 居民电子健康档案的质量现状 R 语言基础 （R 软件安装及运行，R 包安装及调用，基础语法介绍，数据导入导出） AI 工具辅助 R 语言代码编写和代码解读示例 居档合并和匹配 （不同年份、不同区域居档合并，居档与其他来源数据匹配）	佛山
		2	居档质控 （数据查重去重、完整性核查、异常值核查、逻辑检查等）	
		3	居档数据描述 （档案总量、分区域人群档案数量等） 居档质控相关指标的计算 （计算重复率、缺失率、建档率、使用率、覆盖率、体检率、规范率等）	
		4	数据可视化 （绘图基础语法，基于 ggplot2 制作柱形图、箱式图、误差棒图、热力图、分面图等，图形修饰技巧）	

附件 2

2025 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	智能数据分析与可视化技术培训班	1	Python 语言基础 （Python 语言简介、安装及运行、数据类型和结构、基础语法等） AI 辅助 Python 编程	佛山
		2	Python 统计分析相关库概述 （Numpy 库、Pandas 库、Scipy 库和 StatsModels 库概述）	
		3	Python 数据可视化方法与医学实践 （Python 绘图基础，Matplotlib 库和 Seaborn 库概述，绘制散点图、折线图、饼图、箱式图、玫瑰图、聚类热力图等）	
		4	Python 数据统计分析方法（1）： 1. Python 统计分析步骤及技术路线 2. 数据清理 3. 数据转化 4. 统计描述（计数资料描述、计量资料描述）	
		5	Python 数据统计分析方法（2）： 1. 假设检验（分布检验、方差齐性检验、t 检验、Mann-Whitney U 检验、卡方检验和方差分析，以体检数据、疾病监测数据、医学试验数据、实验检测数据分析为例） 2. 分布拟合分析（正态分布、二项分布和 Poisson 分布拟合，以体检数据分析、药物疗效分析和新生儿死亡水平评价为例）	
		6	Python 数据统计分析方法（3）： 1. 相关分析（Pearson、Spearman、Kendall 相关系数及相关矩阵，以体检数据分析为例） 2. 回归分析（一元线性回归、Logistic 回归、混合线性回归，以体检数据分析为例）	
8 月	卫生健康机器学习方法与实践技术培训班	1	Python 语言基础 （Python 软件安装及运行，数据类型结构，基础语法介绍） AI 辅助 Python 编程 Python 语言可视化技术基础 （Python 绘图基础和 Matplotlib 库绘图技术介绍） 机器学习概述 （机器学习方法概述及其在卫生健康领域的应用）	珠海
		2	机器学习 Scikit-learn 库概述 数据预处理方法 模型评估和模型选择方法	
		3	机器学习分类方法与实践： 1. Logistic 回归二分类和多分类（以体质分类预测为例） 2. 朴素贝叶斯分类（以体质分类预测为例） 3. 随机森林分类（以血压分类预测高血压为例） 4. 支持向量机分类（以体质分类预测肥胖为例）	
		4	机器学习回归方法与实践 （以体质指数预测为例）：	

			1.线性回归 2.多项式回归 3.岭回归 4.最近邻回归	
		5	机器学习聚类方法与实践： 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） 机器学习降维方法与实践： 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介绍 2.图像数字化处理 （以 5 种食材照片的数字化处理为例） 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） （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）	
11 月	智能数据分析与预测高级技术培训班	1	Python 语言及统计分析基础 （Python 软件安装及运行，数据类型结构，基础语法介绍，数据导入导出，统计分析相关库概述） AI 辅助 Python 编程 Python 语言可视化技术基础 （Python 绘图基础和 Matplotlib 库绘图技术介绍）	广州
		2	回归分析和生存分析： 1.Logistic 回归（以健康体检数据进行体质类别分析为例） 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）	
		3	判别分析和聚类分析： 1.判别分析（Fisher 判别、Bayes 判别，以体质分类、疾病判别分析为例） 2.聚类分析（层次聚类、K-means 聚类，以体检数据聚类分析为例）	
		4	主成分分析和因子分析： 1.主成分分析（以复方药物化合物的主成分分析为例） 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） 3.因子分析（以体检指标数据进行因子分析为例）	
		5	时间序列趋势分析与预测： 1.ARIMA 模型（以手足口病月发病数变化趋势分析预测为例） 2.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) 3.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)	
		6	传染病动力学模型分析 (基本再生数、实时再生数、SEIR 模型、 SEEIR 模型、SEEIR 模型分析, 以传染病流行特征及趋势分析预 测为例)	

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，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。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）。

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，将于 2025 年 3 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 40 余万字。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、最近邻回归、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、Meanshift 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、线性判别降维、因子分析和使用特征选择进行降维 8 种。本书的案例实践中，还应用了相关 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。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）。