

《R 语言数据分析》常见问答

Q1. 这门课面向哪些人群？

A1. 高年级本科生及低年级研究生。

对机器学习、数据挖掘、数据分析感兴趣的工程技术人员。

涉及的专业包括但不限于人工智能、智能科学与技术、计算机科学与技术、统计学、生物信息学、自动化、经济管理等。实际上，随着数据时代的来临，数据分析几乎渗透到每一个专业领域。

Q2. 这门课程的定位是什么？

A2. 数据科学领域的入门课。

Q3. 学好这门课之后，可以独立从事数据科学相关的工作么？胜任数据科学家/数据工程师等相关职位么？

A3. 应该不可以（假如仅仅看一遍视频的话，那肯定不可以）。

要从事数据科学相关的工作，开展数据科学相关项目，或是谋求一份数据科学家、数据工程师相关的职位，应该不存在速成的方法。这个领域门槛并不低，既要懂编程、更要懂算法，尤其是需要业务敏锐性，能够把实际的问题转换为数据科学问题加以解决。要实现这样的目标，并不是一两门课可以解决的，更多应该是大量开展实际的项目，在实际的项目中提升自己。

这门课的作用在于让大家入门机器学习/数据挖掘，在方法论层面抓住其本质、在工具层面掌握 R 语言、在技术层面掌握典型算法模型的基本原理及其实现。应该说，通过这门课来入门，应该是没问题。至于后续发展如何，那就是师傅领进门、修行在个人了。

Q4. 学习这门课程，需要那些基础？

A4. 需要有大学计算机及数学基础。

所谓计算机基础方面，是指对编写代码有基本概念，接触过一门高级语言编程最好。

所谓大学数学基础，主要是指学过微积分、线性代数、概率论与数理统计，或是相近的课程即可。可能对于某些非理工科专业而言，数学方面的要求会低一点。不过影响也不大。课程里边对于相关的数学知识，也是尽量采用通俗的语言进行引入。

以上内容不一定要记得多清楚，多多少少有点印象即可。

Q5. 学习这门课，需要学编程么？

A5. 需要。

采用的编程语言是 R 语言。

当然，这里的编程主要是指科学计算、统计编程，有别于一般的应用程序的编写。所以，并不会要求在软件编写方面多么训练有素，也不会要求具有软件工程之类的基础，甚至对于数据结构、数据库原理等，也不会有太多特别的要求。

Q6. 不想编程，那这门课还能学么？

A6. 假如是开展数据分析的学习，或是从事相关的工作，还是了解一点编程比较好。傻瓜式的数据分析、数据挖掘软件也很多，但是建议自己写代码。

如果实在对编程不感兴趣，可以看一看《上部：论道》部分。这一部分，主要是方法论和数据分析的核心理念。对于某些对技术不感兴趣的小伙伴，看一看这部分内容，应该也有所裨益吧。

Q7. 学习这门课程，需要有 R 语言的基础么？

A7. 不需要。

课程内容是相对自洽的。即便从未接触过 R 语言，学习起来应该也没有任何压力。当然，假如之前自学过 R 语言，上手会更快一点。即便是学习过其他 R 语言课程，学习本课程《中部：执具》时，应该对 R 语言也会有新的认识。

Q8. 学习这门课程，需要机器学习/数据挖掘的知识么？

A8. 不需要。

这门课的定位是入门课。所以并不假定同学们掌握了机器学习/数据挖掘的相关知识。即便是零基础，应该也可以学习。

Q9. 完整的学习这门课，需要多长时间？

A9. 视频时间约 950 分钟。

换句话说，假如只是单纯的看视频的话，也就是 $950/60 \approx 16$ 个学时。不过，推荐的学习时间，应该会远高于 16 个小时。个人觉得 R 语言、数据分析的学习，并不是多少分钟、或者是十几个小时就能学会的。能照着别人的代码，跑一遍出个结果，这并不代表掌握了 R 语言，更不代表理解了数据挖掘/机器学习。一般来说，两个月左右，应该能上手吧。R 语言本身是易学难精，真正想精通 R 语言数据分析，需假以时日。我自己讲这门课，目前应该是第 13 轮吧，我对我自己的定位是 R 语言学习者和爱好者，精通二字，还遥不可及。不过并不影响我利用它开展学术、科研及具体的工程项目。

Q10. 什么时候开始上手写代码，是在看完《上部：问道》之后么？

A10. 课程的内容体系是先讲总的方法论、然后讲 R 这个工具。有了方法论和 R 这个具体的实现工具，最后讲具体的算法模型。

不过，在学习的过程中，建议一边看视频、一边练习代码。也就是说，在学习《上部：问道》的同时，就可以开始尝试写代码。写代码是一件很有意思的事情，早点开始，早点获得其中的快感和成就感。

Q11. 这门课有配套教材么？

A11. 没有。

这门课的筹备，应该是在 2010~2011 年左右，2012 年正式开课。那个时候 R 语言的教材还相对较少，课程的大部分内容，都是任课老师自行编制（当然除了自身的项目积累之外，也参考了大量网络资源）。在每一轮授课的过程中（秋季课程、暑期课程），都会根据自身的积累和体会对授课内容进行优化和调整。所以课程内容并非取自某本教材或专著，自然也就没有配套教材。

如果顺利的话，机械工业出版社将于 2021 年出版授课老师编写的《机器学习赏析》教材，内容与本课程基本一致，但有较多扩充和完善。

Q12. 这门课程适合作为翻转课堂的学习资源么？

A12. 应该适合。

具体问题请直接与授课老师取得联系（axb@bupt.edu.cn），将尽力提供协助。

Q13. 课程配套的代码和数据，在哪里下载？

A13. 课程全部代码及数据，托管在 Github，请自行前往下载。

地址：<https://github.com/byaxb/RDataAnalytics>

如果下载不了，可通过永久邮箱地址 axb@bupt.edu.cn 索取。

Q14. 课程的讲义提供下载么？

A14. 暂不提供。

Q15. 这门课可以通过哪些平台学习？

A15. 目前本课程仅在“学堂在线”对公众开放。

Q16. 这门课有相应的 QQ 群或微信群么？

A16. 已建立 QQ 群，群号为：795827473。

Q17. 若发现课程讲授内容有错漏，通过何种渠道告知授课老师？

A17. 限于老师本人的水平有限（和时间没关系哈，毕竟任何人都是在有限的时间里开展工作），错漏之处肯定不少。

若发现有问題，请在讨论区发帖留言，讨论区由授课老师及助教维护，大家的留言都能看到，对于相关问题我们也会及时解答；也可以通过 axb@bupt.edu.cn，将相关问题告知授课老师。

Q18. 这门课程的成绩是如何计算的？

A18. 课程成绩由四部分构成：

视频单元考核（20 分）：对于每一段视频，观看超过 90%得满分，否则不得分

讨论单元考核（10 分）：在讨论区的课程讨论题《如何发挥人工智能的头雁效应》主题下有效回帖即可得分

作业单元考核（40 分）： $40 * (\text{学生作业得分} / \text{作业总分})$

考试单元考核（30 分）： $30 * (\text{学生试卷得分} / \text{试卷总分})$

注意：按照学堂在线的计分规则，讨论单元是否得分，取决于是否在讨论区《如何发挥人工智能的头雁效应》这一特定主题下有效回帖。并不是在讨论区发言或回帖均得分。

Q19. 如何获得这门课的认可证书？

A19. 在“学堂在线”开展学习，分为“免费学习”和“认证学习”两种方式。只有“认证学习”才提供认证证书。

Q20. 免费学习和认证学习的课程内容有区别么？

A20. 没有区别。

Q21. 学习完这门课之后，接下来学习什么？

A21. 取决于你自己的学习兴趣和研究兴趣。

入门机器学习/数据挖掘之后，会发现这是一片广阔的天地。比如，自然语言处理、医学图像处理、空间数据分析、时间序列、生存分析、生态环境数据分析、复杂网络分析、计量经济学、金融、……。应该说，不同领域，R 都有很好的支持。

大家可以看一看 TASK VIEWS 中是否有相同或相近的主题，或是通过 Github 等其他渠道，找一些自己感兴趣的主题。相信假以时日，能成为相关领域的数据分析专家。