

matplotlib 课外练习

请编写代码：

构建一个数据模型，由电信科专业20级五个自然班，每班人数45，数据包括学生的“学号”、“姓名”、“班级”以及“高等数学”、“英语”、“Python”、“普通物理”、“科学导论”五门课程的成绩，成绩数据的产生要符合正太分布，每门课最高分不超过100，高等数学平均分为70，英语平均分为85，Python平均分为80，普通物理的平均分为75，科学导论的平均分为88。“学号”由10位数字组成，前六位为“202021”，最后两位为班级学生的序号(01-45)，中间两位表示“班号”(01-05)，分别使用随机函数库(np.random, faker)创建225个学生的数据。使用pandas库将以上数据生成DataFrame类型数据，再指定学号作为行索引，各列数据的次序调整为班级、姓名、高等数学，英语、Python、普通物理、科学导论，然后再分别保存为.csv和.xlsx两个数据文件。

根据以上数据，先进行异常值检查，将超过100分的成绩全部改为100，再使用matplotlib.pyplot库，

- 1、在一张图表上分别绘制出电信科20-1班五门课成绩的拆线图；
- 2、在一张图表上分别绘制出电信科20-2班五门课程平均成绩的柱状图；
- 3、在一张图表上分别绘制出五个班级Python课程成绩的箱式图，比较这五个班级Python课程的成绩分布情况；
- 4、取bins参数值分别为默认值、18、27、36和45五个值，分别绘制电信科20-3班五门课程的直方图，比较bins参数值对直方图的影响，分析这五门课直方图对比的含义；
- 5、对五个班的“普通物理”课程成绩按
“优”： ≥ 90 、“良”： $\geq 80-90$ 、“中”： $\geq 70-80$ 、“及格”： $\geq 60-70$ 、“不及格”： < 60 ，统计计数，再以饼图绘制出各成绩所占人数比例构成。

建议在jupyter notebook中完成，所有绘制的图表，应配图标题，图注，坐标轴标题，标签刻度等相应信息。

保存在以自己“学号+姓名”为名的文件夹中，然后在下列截止日期之前，上传到“第11周课外作业”文件夹中。

本次作业上交截止时间为：**2021年5月23日**，