

Biostatistics (BIOL0031132104) - Assignment #4
(released on Dec. 13th, due on Dec. 24th)

1. 在全球变暖的影响下，许多生活在山地的物种正向海拔更高处移动。那么，这些物种的垂直分布范围（即海拔范围上限与下限之间的距离）是否也在发生变化？研究人员于 2017 年调查了秘鲁潘蒂亚科拉山（Cerro de Pantiacolla）上的鸟类，并将调查结果与 1985 年同一座山的调查进行了比较。数据文件

“chap11q27birdElevation.csv”中“changeExtent”列的数值是 14 种中海拔鸟类垂直分布范围的变化，单位为米。负值表示该物种的垂直分布范围有所缩小，正值表示有所扩大；而 0 表示没有变化。如果我们想要检验山地鸟类垂直分布范围没有变化的零假设，即山地鸟类垂直分布范围的平均变化为 0 米，请依据 *t*-test 给出完整的检验过程，包括零假设和备择假设、样本数据的均值、统计检验量值、P 值、并计算置信区间，给出最后的统计学和生物学结论；同时给出 R 代码。[20 marks]

(a) 零假设和备择假设；[4 marks]

(b) 样本均值（保留小数点后三位）；[2 marks]

(c) *t* 统计量（保留小数点后三位）；[2 marks]

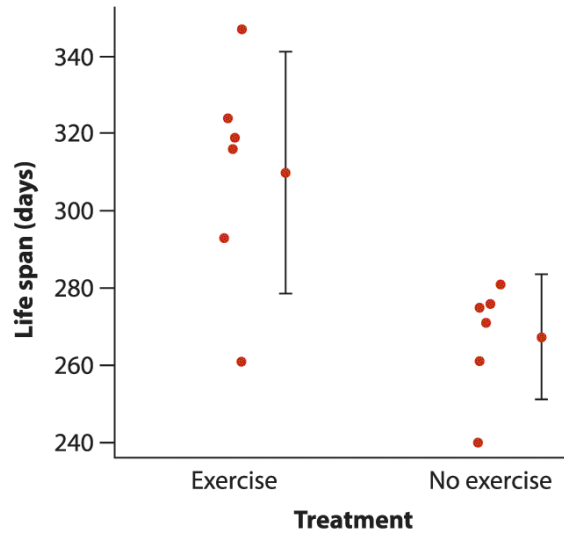
(d) P 值（保留小数点后三位）；[2 marks]

(e) 置信区间（保留小数点后两位）；[2 marks]

(f) 结论；[4 marks]

(g) R 代码；[4 marks]

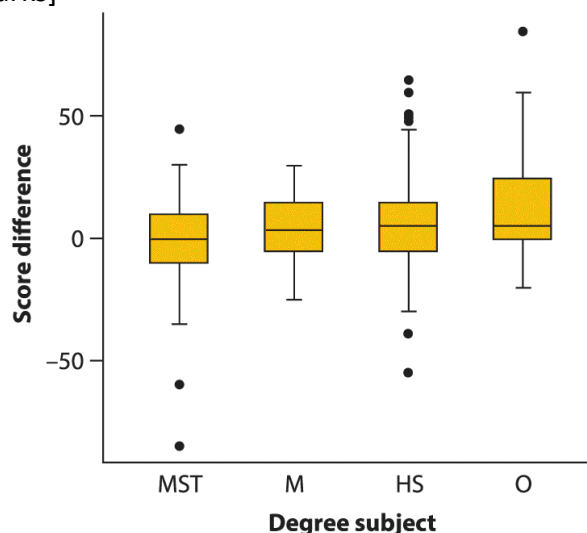
2. 脊髓小脑共济失调 1 型（Spinocerebellar ataxia type 1）是一种神经退行性疾病，其特征是运动能力逐渐丧失，最终导致死亡。该病是由 Ataxin-1 基因编码区中的 CAG 重复序列的扩增（an expanded CAG repeat）引起的。研究人员基于该疾病的小鼠模型研究了运动对治疗这种疾病产生的有益影响的可能性。小鼠模型是在具相同 Ataxin-1 基因的小鼠中“敲入（knocked in）”了一个扩大的 CAG 重复序列，使具有类似的症状。数据文件“chap12q28lifeSpan.csv”中给出了六只接受运动疗法（exercise）的小鼠和六只未接受运动疗法（no_exercise）的小鼠的寿命（天数）。每个处理组数据和 95% 的置信区间见下面的带状图（strip chart）。[30 marks]



Whitlock & Schluter, *The Analysis of Biological Data*,
3e © 2020 W. H. Freeman and Company

- (a) 依据上图，你能否预测“两种治疗方法对患病小鼠平均寿命的影响是否不同”正式统计检验的结果？请解释原因； [4 marks]
- (b) 如果要对“两种处理方法对患病小鼠平均寿命的影响相同与否”进行假设检验，可以选择什么检验方法？ [2 marks]
- (c) 请计算运动组和不运动组各自的平均寿命均值： \bar{Y}_1 和 \bar{Y}_2 （保留小数点后两位）； [2 marks]
- (d) 基于你选择的检验方法，进行正式的统计检验，包括零假设和备择假设； [10 marks]
- (e) 基于合并方差来计算两样本均值差异的标准误 $SE_{\bar{Y}_1 - \bar{Y}_2}$ （保留小数点后三位） [2 marks]；参考公式为：
- $$\text{合并方差: } s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$
- $$\text{均值差异的标准误: } SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$
- (f) 给出置信区间；并基于该区间回答运动对平均寿命的延长天数是多少（保留整数）？ [2 marks]
- (g) 给出 R 代码； [8 marks]

3. 在科研论文中加入数学内容是否会让读者认为论文更有价值呢？研究人员向 200 名拥有研究生学位的人随机发送了两篇科研论文摘要。在每位参与者收到的两篇摘要中，其中一篇的文中插入了一个无意义的描述数学模型的句子，与论文主题没有任何概念上的关联；而另一篇摘要中则没有添加任何数学内容。参与者被要求用 1 分（低分）到 100 分（高分）对每份论文摘要的科研质量进行评分，并记录两份摘要得分的差值：有数学内容的摘要的得分减去无数学内容的摘要的得分。研究人员还记录了参与者的研究生学位的学科（major），并分为四类：MST（包括数学、科学、技术，math, science, technology）；M（医学，medicine）；HS（人文、社会科学，humanities, social science）；或 O（其它，Others）。以下是各学科组数据的箱型图 boxplot 和基本统计量表；完整的数据见文件“chap15q24MathScore.csv”。请回答以下问题：[28 marks]



Whitlock & Schluter, *The Analysis of Biological Data*,
3e © 2020 W. H. Freeman and Company

表 1. 各学科参与者对两篇科研论文的评分差异的均值、标准差及样本大小。

Degree major	Mean score difference	SD score difference	n
MST	-1.28	19.24	69
M	3.06	15.99	16
HS	6.60	21.25	84
O	13.9	23.31	31

- (a) 通过肉眼检查图表，判断数据对 ANOVA（方差分析）假设检验前提（assumptions）的满足程度，符合或可能不符合前提的情况分别有哪些？[8 marks； 每条 2 marks]
- (b) 检验参与者的学科背景是否会影响他们对含有无意义数学陈述的论文摘要的平均看法，包括给出零假设和备择假设、ANOVA 检验表格、以及结论。

(表格如下: 可通过表 1 数据计算; 或从 R 语言读取原始数据后进行 ANOVA 检验获取; 此外, F 临界值为 2.65) [16 marks]

Variance Source	SS	df	MS	F	P
Groups					
Error					

(c) 参与者学科背景与评分差异之间的关系是否紧密? 用 R^2 回答。[4 marks]

(d) 不要求给出 R 代码, 但若给出的话有额外加分。[bonus 5 marks]

4. 一项研究的目的是探讨 "水蛭疗法" 对膝关节骨性关节炎引起的疼痛的治疗效果。两种治疗方法被随机分配给了 51 名膝关节骨关节炎患者。第一种是水蛭疗法, 接受该种疗法的患者一次性在患病膝关节的软组织上放置 4-6 只药用水蛭, 让水蛭自由进食, 直到它们自行脱落 (平均 70 分钟后)。第二种为对照组, 患者被告知每天使用双氯芬酸凝胶在患处涂抹两次。由不了解每位患者所接受治疗的人员对患者进行问卷来评估疼痛。结果显示, 治疗开始七天后, 水蛭疗法组的疼痛明显减轻。[22 marks]

(a) 这项研究是否包括对照组? 请解释。[5 marks]

(b) 这是实验研究还是观察研究? 请解释。[5 marks]

(c) 这是随机试验设计吗? 是否包含区组/组别 (block) 设计? (如果不包含区组的话则为完全随机设计)。请解释。[6 marks]

(d) 本研究没有采用哪种减少偏差 (bias) 的策略? 不采用该策略会对结果产生什么影响? [6 marks]