

“语言、数据与研究”QQ 群互动问答操作指引

1. “语言、数据与研究”系列课程包括《阿檀小倪讲量化》、《量化工具百宝箱》和《量化研究论文写作》。欢迎学员就上述课程学习相关的议题进行沟通交流；
2. 为确保每位学员的问题都能被及时处理，该 QQ 群设有两名微课助教：郭凯同学和孙望同学。在周一至周五的正常工作时间内，微课助教会及时汇总学员们的的问题；
3. 微课助教在汇总学员问题后，会对问题进行初步分类。经分类后的问题，会分别由金檀博士、倪竞博士或李百川博士进行回答。回答会及时发布在 QQ 群或私信回复相关学员；
4. 微课助教定期对问答进行汇总和筛选，经匿名化处理后形成《阿檀小倪讲量化：常见问答札记》。常见问答札记定期发布和更新（学员不同意收录的问题除外），供 QQ 群中的学员参考；
5. 由于时间和人力限制，与课程学习内容没有直接关联的问题，课程团队会根据问题本身的性质进行妥善处理。敬请谅解。

感谢您的支持，希望“语言、数据与研究”能与您共同成长！

“语言、数据与研究”课程团队

2016 年 12 月（更新）

《语言、数据与研究：阿檀小倪讲量化》

常用问答札记

(2016年5月—12月问题精选)

主讲人：金檀、倪竞

数据顾问：李百川 课程助教：郭凯、孙望

第1讲 研究：用数据回答问题

问题 1:

其实几位老师讲得都很好，条理很清晰，通过这个平台，我们接触到很新的研究工具，也明白了一些研究方法。可是老师有没有觉得，如果连国内的外语核心期刊都望尘莫及，国外期刊这些大型的文本研究是不是更遥远？

回答:

在确定论文案例时，我们主要考虑该案例对量化研究方法的示例作用。我们最初也选了2篇中文论文作为案例。但在试讲过程中，发现这2篇文章由于篇幅较短（8-10页），对研究方法部分的描述较为精炼。我们希望论文案例对研究方法能有详细的“示例”作用，所以这2篇文章最终没有选用。其实无论国内还是国外期刊，都有很多的精品之作。由于时间问题，这一版本我们只收录了5篇英文论文案例。我们还在陆续补充这门课的论文案例库，其中也有很多近期发表的中文论文，希望后续能有机会与大家分享。另，这门课主要关注研究方法的讲解，我们近期也在考虑论文写作课程的设计。

问题 2:

等距变量和等比变量的区别能否再详细解释下，为什么等距变量如温度没有零点，不能乘除？

回答:

两者的差距就是有没有绝对零点。温度是有绝对零点的，是-200多度，但是摄氏温度没有绝对零点，还会有零下多少度之类的存在。不能乘除的意思是，20度不能说是10度的两倍那么热，50度也不是10度的5倍那么热。

问题 3:

口红的颜色在变量视图中的度量标准为什么选择度量，而非名义？

回答:

我们可以用红色的深浅从1到10来评判口红的颜色，这样可能对后续的数据分析有更大的可选择性。当然，口红颜色也可以是名义变量，那后面就有很多数据分析没办法做了，因为不满足基本前提假设。

问题 4:

检验问卷质量可以用 SPSS 进行信度和效度检验，但据说数据呈正态是使用 SPSS 进行各种统计分析的前提，请问是这样吗？

回答:

确实，数据的正态性是很多统计分析的前提之一，但是如果你的样本量太小的话，是很难得到正态分布的样本的。

问题 5:

实证研究的样本量有什么要求？

回答:

根据研究问题不同，具体数量要求不同。但是，一般来说样本量要大于 30。

问题 6:

假设我预备给 200 名学生做一个问卷，实测前先选 20 名同学做 pilot study 以评价问卷质量，请问是否需要先对 20 名同学的样本进行正态分布检验？假如检验结果不呈正态，怎么办？

回答:

20 名同学的样本太小，非正态的可能性很大。但是检验结果的正态性，与问卷质量没有什么联系。要对问卷进行评价，需要其他手段。

问题 7:

我想统计一下学生作文使用语块的情况，请问使用什么工具好？需要我自己标注文本吗？

回答:

具体的计算应根据具体研究情景和目的而定。无论机器协助亦或人工标定，建议都对适用于自己研究的“语块”进行操作性界定。2015 年有一篇相关论文（包括这篇论文的参考文献），可以参考一下：AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, 17, 51-62. 另，考虑到很多老师对应用语言学领域中的工具使用感兴趣，我们正在研发相关工具课程，后续会有具体信息发布，敬请关注。

问题 8:

如何将多个维度的调查数据整合成一个变量，比如，我做学习动机研究，根据定义，学习动机是有内在动机，在外动机，学习意愿这三个维度构成，设计问卷时也是从这三个维度进行调查的，现在就是不知道怎么把这个维度的数据整合成学习动机变量，进行后面的学习动机水平与学习策略的相关性研究。如果能构成一个潜在变量，怎么把这三个维度的数据整合成一个变量的数据呢？

回答:

如果已经有了理论构想，那么在收回数据之后，可以进行验证性因素分析，看这三个测量维度是否可以组成一个潜在变量。如果构成一个潜在变量，可以直接在结构方程模型中使用该潜在变量，或者根据三个维度在潜变量上的载荷作为加权，计算一个分数。前者更好。

问题 9:

写大学生学习动机与二语习得关系的论文，一般测量量表用什么？在哪获得？样本多大合适？

回答:

具体的测量量表请参考相关的文献。量表可以通过硕博士论文或联系原作者取得，通常论文的原作者会很乐意分享，但是请务必记得获得作者的同意。样本的大小要看研究的复杂程度和研究目的，如果只是两个变量的话，理论上 30 个被试就可以进行相关分析。可以考虑进行 power analysis 来确定样本量。

问题 10:

写“母语负迁移在大学生汉英翻译中的体现和成因分析”的论文，样本多大比较科学？

回答:

样本的大小要看研究的复杂程度和研究目的，并没有说有一个“科学”的样本量。可以考虑进行 power analysis 来确定样本量。

问题 11:

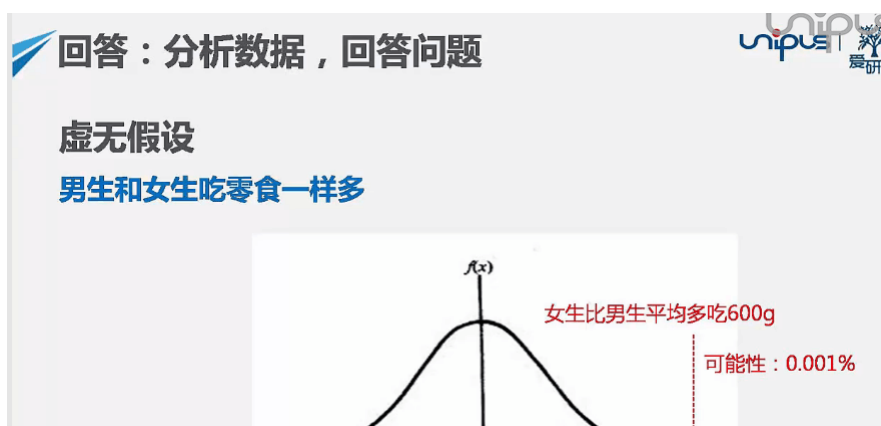
有专业性和权威性强一些的词汇量测试工具推荐吗？

回答:

向您推荐 **Vocabulary Levels Test** 这个词汇测试工具。

问题 12:

请问一下，在这张图里边，女生吃的比男生多的几率是 0.001，这不是说明女生比男生吃的多不可能发生吗？既然女生不比男生吃的多，那就有两个可能，要么一样多，要么男生吃的更多是吗？这里是怎么证明上面假设不成立的？

**回答:**

这张图里的假设是“女生和男生吃的一样多”，现在的数据则落到了 0.001 处，说明这件事是不可能发生的，所以女生吃的比男生多。

第 2 讲 描述：大致情况如何？

问题 1:

在 Stuard Webb 的文章中 28 万字三个语料库是如何制作的？

回答:

研究中的主要用途是通过将文本电子化，然后使用词汇分析工具来实现的。如果只是一般的语料用途，过程可以参考我们“试题语料库”的建设过程。详情可参阅：

<http://languagedata.net/exam/>

问题 2:

课程中提到的 range 软件，网页中有几个版本，应该下载哪个呢？

回答:

这些版本是基于不同词表对文本进行分析的，您可以根据自己的需要下载不同版本。它们给出的数据结果类型都相同，只是基于不同词表对文本进行分析。

问题 3:

如何将 Range 工具的结果数据导入 SPSS 软件？

回答:

由于 Range 工具生成的数据结果为 txt 文本格式，因此，需手动将数据粘贴到 SPSS 软件中。可以将 txt 中的结果先粘贴到 excel 表格中，这样，就可以单独选中 Frequency 数据，然后，就可以直接粘贴到 SPSS 中。将数据结果粘贴到 excel 后，也可在 SPSS 软件中进行导入，选择相应的表格范围即可，但工作量和刚刚提到的粘贴差不多。

问题 4:

Range 工具生成的为单个单词的词频，而非累积词频的问题，应该怎么办？

回答:

可以将结果数据复制到 SPSS 软件中，然后按照 SPSS 的流程进行计算即可。

问题 5:

打开 Range 软件后，具体怎么操作？能否提供更清晰的指导。

回答:

关于 range 工具的操作，您可以参考 PDF “Range Instruction”。关于工具的使用和操作，我们正在研发一门新课“iMethod 工具百宝箱”，主要讲授国内外常用的词汇、句法和篇章分析工具，其中也会包括 range 工具，到时候也希望您关注。

问题 6:

Stuard Webb 的文章中 28 万字是不是三个语料库制作成 txt 格式的文件，然后直接可以用 Range 软件来分析？

回答:

在使用 Range 工具时，所分析的文件需是 txt 文件格式。所以，只要是 txt 文件，就没问题啦。您可以到我们提供的网站上下载 Range 工具试用一下。

问题 7:

在第二讲中，小倪老师对 SPSS 的操作讲得很清楚，但关于其输入的数据（词汇类别）的性质，我还是不清楚。这是 Range 处理文本的结果吗？

回答: 这是我们根据研究问题，使用 Range 软件进行分类的结果。如果一个词属于 1000 以内的常用词，就标记其词汇类别为“1”。然后，我们接着统计总计 855429 个词汇中，有多少个 1，多少个 2，多少个 3……因此使用了频率统计。

问题 8:

第二讲 Webb&Macalister(2013)论文中，table3-6 里面的数据是否是将儿童读物、分级读物和大人读物这三类语料用 Range 分析出来的吗？为什么小倪老师展示的是用 spss 进行每个单词所属等级的频率统计？论文中每个文本库都含有 28 万词，每个单词所属等级是怎么录入 spss 的？

回答: 这些词是用 range 进行分类的，然后复制粘贴到 SPSS，用 SPSS 统计频率。

问题 9:

请问 range 词汇测试工具能测学生的口语词汇量吗？

回答:

正如第二讲中所述，Range 工具估算的一般为输入型文本的词汇需求量，较为典型的为阅读文本。对于输出型词汇量的估算，一般不能使用 Range。关于输出型词汇量的研究，这是一个非常有价值的议题。Edwards 和 Collins 在 2011 年发表了一篇文章讨论写作词汇量的估算问题，颇具新意，可以参考一下：Edwards, R., & Collins, L. (2011). Lexical frequency profiles and Zipf's law. *Language Learning*, 61(1), 1-30. 然而，到目前为止，据我所知，输出型词汇量，尤其是口语词汇量的估算，还没有很好的解决方案。

问题 10:

关于第二讲论文，我自己用 Range 做了一个分析，得出这样的结果：

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	3231/81.28	752/58.98	516
two	265/ 6.67	181/14.20	163
three	105/ 2.64	82/ 6.43	5
not in the lists	374/ 9.41	260/20.39	?????
Total	3975	1275	754

问题 a. 在我分析的这个文本中，81.28%的单词属于第一级别，即最常用的 1000 个单词。6.67%的单词属于第二级别，即 1000-2000 个最常用单词。2000 词的 Cumulative Coverage

是 $81.28\%+6.67\%=87.95\%$ ，3000 词的累积覆盖率是 $81.28\%+6.67\%+2.64\%=90.59\%$ 。也就是说最常用的 3000 词占这篇文本总词量的 90.50%。请问老师，这样解读数据是否正确？

回答：

没错，您的解读是正确的。

问题 b. 所谓 token 就是文本中的单词数量？而在计算 type 时，同一个单词只计算一次？第二讲论文中 Table 3-5 的中的 Cumulative coverage 数据应该都是 token 的数据吧？

回答：

是的，您对 word token 和 word type 的理解正确，论文中的数据都是关于 token 的数据。

问题 c. 为什么我的数据中没有 Proper nouns 和 Marginal words 这一项的统计？事实上，我的文本中包含有专有名词。

回答：

Range 是以不同词表为基础对文本进行分析的。分析结果中的 *not in the lists* 表示没有出现在词表中的单词。由于大多数词表不包含 Proper nouns 和 Marginal words，所以您可以在 *not in the lists* 中进行人工校对。但如果您要找的 Proper nouns 和 Marginal words 已经包含在词表中，则不会出现在 *not in the lists* 中。

第3讲 相关：它俩有何关系？

问题 1:

这个研究问题应该讨论的不是相关，而是差异吧。各个语料库有多少重叠，也就是有多少相同，有多少不同吧？为什么是相关呢？两个不同的事物会相关。这是同样都是核心通信词汇表，只不过基于不同的语料库而已。应该看差异。用独立样本 t 检验吧。

回答:

相关检验的是两个变量之间的关系。在本案例中，因为比较的是语料库之间的联系，所以联系越紧密，我们越可以认为这两个语料库的相似性就越高。因此，原作者想通过调查词序之间的相关来比对语料库之间的重叠，用相关并没有错，而且这是一种比较聪明的角度。而 t 检验很难比较两个语料库的差异，因为 t 检验比较的是平均数的差异。

问题 2:

两两比较词表，那如果有 A 词表里的词没有出现在 B 词表该怎么标注顺序号呢？小倪老师的例子里，似乎 A 词表和 B 词表的词是一样的，只是顺序不一样。

回答:

这种情况下，A 词表对应的这个词，在 B 词表的排序就是空白的（或者编为 0），而这种排序方式并不影响相关结果。我举一个极端的例子：假设 A 词表中的排序的第一个词，就不存在 B 词表中，而其他顺序都一样。那么 A 和 B 的相关也仍然是 1。

问题 3:

样本量不足 30 的非正态分布数据如何观察两两相关？

回答:

只要总体正态，就可用同样方法计算。但是样本量太小，结果不可靠。

问题 4:

在 SPSS 中录入数据时，第一个语料库的录入只要将数字 1 到 3000 依次从上到下录入。录入其他三个语料库数据时，需要首先找到一个单词在第一个语料库中的排位，然后查找该单词在第二、三、四个语料库中的排位。假设 cat 这个单词在四个语料库中的排位分别是 80,85,90,95。那么在 spss 表格的第 80 行的数值即：80,85,90,95。请问老师：我的理解是否正确？如果这样录入，工作量非常巨大，研究者就是这样人工录入吗，还是或许有什么简便的方法？

回答:

理解没错。可能有其他方法或工具，但是论文中没有特别详述，所以我们也不完全清楚。建议联系原作者。

问题 5:

假设 scandal 这个单词在第一个语料库中排位 2900，但是没有出现在后三个语料库的最常用 3000 词中，那么后三个语料库对应 scandal 的格子怎么录入？

回答：

虽然作者没有明说，但是应该是空白，作为缺失。

问题 6：

后三个语料库中必定有部分单词没有出现在第一个语料库的最常用 3000 词中，这些单词怎么录入？是否在第 3001 行开始录入？那么最后的表格应该是第一个语料库有 3000 个数据，但第二、三、四个语料库的数据会超过 3000，也就是说 4 个语料库所包含的数据量是不等的。由此我想到一个问题，进行两两相关比较时，进行比较的两组数据所包含的数据量是否要求是一致的？还是可以是不一样的？

回答：

数据量是相等的，第二个语料库中有的词汇，第一个没有，那么第一个语料库中就会出现空白数据。参见上一问。两组数据的数据量一定是相等的，空白的数据也是数据。

第 4 讲 回归：它们有何关系？**问题 1:**

在第四讲中关于提炼核心概念部分，是将影响学生口语成绩的语音、流利度、词汇、语法四个因素提炼为区别性特征。请问这个区别性特征，是你们参考其他文献总结出来的，还是你们已学知识中就已经知道的呢？个人理解，提炼核心概念部分是从实践提升到理论的一个过程，如果单从教学实践中找到情景不难，但是如何转换为核心概念不容易。对于我们理论知识没有那么系统的情况下，应该如何做呢？

回答:

“区别性特征”这个概念，是通过文献综述提炼出来的，即“区别性特征”在口语评分中的作用。在汉语作为第二语言的口语评分中，区别性特征究竟选哪些特征呢？这个是通过对比汉语二语评分文献和能力标准的梳理、归纳得出的，论文中有一个表格，您可以参考一下（就是后面用打钩来表示的）。所以，这里有两个步骤，一是核心概念的提炼，二是核心概念的具体操作。一般来说，都是先找到一篇重点文献，锁定一个概念，而后按图索骥，拓展文献阅读量，同时反思总结与实践结合。这个过程也就是形成自己研究问题的过程。

问题 2:

能否把第四讲论文里的研究复制研究翻译课程分数呢？

回答: 可以尝试类似的研究设计。对翻译能力的评估和评分研究，应先界定翻译能力概念及区别性特征的量化等问题。

问题 3:

关于第四讲论文，请问老师进行回归分析时，假如自变量和因变量的统计尺度不一样，比如自变量用的是 Likert 5 点量表统计的数据，而因变量是 100 分制的成绩，是否可以进行分析？还是需要首先将 100 分制转换成 5 分制？

回答: 不用转换，可以直接回归，知识在解释的时候要特别小心尺度的问题。

问题 4:

请教一个问题，在第四讲的习题中有一题我不太明白：

11 在本节的必读论文中，阿檀他们最应当关注的是回归分析结果表中的 _____

2.0分

- A. R方
- B. 标准化系数
- C. F值
- D. F值的显著性

正确答案是 B，你的答案是 A。回答错误

如果最应该关注的是标准化系数，那么也应该同时选择 F 值显著性和 F 值，因为标准化系数、t 值和 sig 三者同时决定了自变量对因变量是否有显著影响，也就是说这题应该选三个答案，不知道我这样理解对不对？

回答：

标准化系数代表了自变量对因变量的影响大小和方向，而 t 值的显著性则代表这种影响是否达到显著水平，R 方检验了所有自变量的总影响之和的大小，F 值检验了所有自变量的总影响之和的显著性，对于多个自变量的回归分析来说通常都是显著的，因此 F 值及其显著性不用特别关注。所以，如果要选一个最应该关注的，应该是标准化系数。

问题 5：

汉语文本指难针中“难度 LD 值”是文本整体难度吗？包括词汇、句法等多个维度吗？这是怎么计算出来的？因为如果要在论文中使用您的这个工具，匿名评审可能会问这样的问题，不知道您是否有难度计算的说明？

回答：

是的，难度 LD 值是文本的整体难度指标。这个指标的计算，在“汉语文本指难针”主页的“视频说明”和“使用指引”中都有简要说明。这里作简要说明（后续我们也会有论文说明）：

- (1) 难度指标计算特征包括：字词等级、平均句长和文章长度等，难度参考基准为“汉语教材语料库”中典型等级的教材文本（languagedata.net/corpus/）。
- (2) 采用机器学习中重要的支持向量机算法。经实验检验，该算法的分级准确率已超过 90%，参见林星彤同学的毕业论文《国际汉语阅读文本难度指针的设计与实现》（中山大学移动互联网与金融大数据实验室）。
- (3) 我们“指难针”的设计理念是：除了文本难度定级外，可提供文本改编反馈。因此，未采用“间接”文本指标，如：语义、连贯等。原计划也采用句法指标，但汉语句法的量化指标尚未成熟，对文本难度的预估效果也不太理想，因此没有采用。多维度汉语特征的计算，如果需要，可参考文献（Sung, Y. T., Chang, T. H., Lin, W. C., Hsieh, K. S., & Chang, K. E. (2015). Crie: an automated analyzer for chinese texts. *Behavior Research Methods*, 1-14.）。

问题 6：

您在讲解论文时说，Lexile 可以测 lexile measure（篇章可读性计算值），请问这个“篇章可读性计算值”对应于您的“难度 LD 值”吗？谢谢！

回答：

稍有不同。我们的“难度 LD 值”，主要指用于教学用途的“文本难度”，因为我们的数据基础主要是百万级汉语二语教材语料库中的教材文本，而 Lexile 的“难度”主要是针对母语阅读“篇章可读性”文本难度估算。关于 Lexile 的相关介绍，可以参考我们微课中推荐的 Lexile 网站信息。

问题 7：

Coh-Metrix 和 Lexile 都可以对文本难度和可读性进行分析，那么两个工具在文本难度和可读性分析的功能上有何差异？

回答：

Coh-Metrix 与 Lexile 的研发背景与目的用途都不相同。前者是基于对文本的不同特征进行量化计算来测量文本的 **coherence**，后者是用于测定文本的难度等级，将不同难度的文本与不同学习阶段学生的阅读能力进行匹配。因此，两个工具设定的测量指标是不同的。

第 5 讲 对比：它俩有何差异？

问题 1:

能否麻烦讲下配对样本 t 检验，和独立样本 t 检验，什么情况下使用，能否举个例？

回答:

配对样本是两个相关联的样本，而独立样本是两个相互独立的样本。比如，如果我们比较男生和女生的数学成绩，就是独立样本；而如果我们比较同一个班的男生其中考试和期末考试之间的差别，就是配对样本。

问题 2:

想知道一个班级的平时成绩与考级成绩之间是否有联系，用什么测量？独立样本 T 检验吗？

回答：建议使用配对样本 t 检验。

问题 3:

t 检验是不是只能体现两组样本的差异，无法得出说哪个教材更好，更适合学生？

回答:

如果你的研究问题，就是通过某个指标来比较哪个教材更好（比如词汇难度）或者更适合学生（比如可读性），那么 t 检验的结果就能回答你的研究问题；如果你只是随便选了两个指标，那么 t 检验当然不能得出结论。

问题 4:

我有两个班做对比实验，其中一个班使用了新的教学策略，另外一个班没有使用，这两个班实验前做了测试，实验中做了测试，实验后又进行了测试，我想知道使用了新策略的班级测试成绩跟对比班级是否存在显著性差异？是应该做独立样本 T 检验吗？

回答:

可以的，注意其他控制因素的说明。

问题 5:

请问如果每组样本量远远超过 30，比如每组为 60 人，能否进行 t 检验和方差分析？

回答:

可以，但样本量大小只是一个因素，还需满足其他前提条件。

问题 6:

请问做二个教学班比较，最小有效样本的对照班级人数和试验班级人数为多少？

回答:

根据研究问题，研究方法等因素的不同，有效样本的数量是不一样的，并不存在统一的所谓最少人数。如果一定要有个数值的话，学术界习惯以 30 人为小样本的分界线，但再强调一句，这个 30 并不是在所有情况下都适用的。

问题 7:

关于第五讲论文，这篇论文使用 t 检验，所有 t 检验的自由度都包含了两个数值，这是为什么？以为我看到的其他论文中汇报 t 值都只有一个自由度。

回答:

这应该是笔误。其实是借鉴了 F 检验的写法，如果你仔细观察，会发现两个自由度的第一个永远是 1，这第一个自由度代表的是分组。但事实上无此必要，因为 t 检验永远是两组比较。

问题 8:

Coh-metrics 和 Lexille 都是文本难度分析工具，它们的区别在哪里？有多少主流的文本难度分析工具呢？怎么根据判断哪个最适合自己的研究呢？

回答:

这两个工具最初的研发目的有些不一样。Coh-metrics 研发最初是从文本“coherence”的角度出发测量文本指标的，所以您可以发现，这个工具的名称就叫 Coh-metrics。Coh-metrics 报道了百余个不多的量化指标。Lexille 则主要是从词汇、句子和篇章长度来评价文本难度。关于工具的使用，我们正在研发一门新课“iMethod 工具百宝箱”，到时会讲授国内外常用的词汇、句法和篇章分析工具。到时候会针对具体的情景来介绍工具，这样您可基于使用情景来判定使用哪种工具了。

问题 9:

Table3、4 和 5 中，M、SD 下方都另有一行，MR、SR，不知道这两个缩写是什么意思？

回答:

表格下面自带说明：Note. MR = mean of ranks; SR = sum of ranks.

问题 10:

第五讲中提到独立样本 T 检验中的样本一般是小于 30 的，那么大于 30 的样本使用什么样的检验方法呢？

回答:

t 检验可以适用于小于 30 的样本量的情况，但是大于 30 也可以使用。当 t 越来越大时，t 检验就越来越接近于 Z 检验。或者，如果样本量远超 30，也可以考虑使用 F 检验。

问题 11:

请问 t 值与 sig.有何区别？

回答:

t 值是 t 检验中计算出来的值，sig.是 t 值对应的概率值，是通过 t 分布所计算出来的。所以 t 值和 sig.不是一个东西，但是在如果自由度固定的话，t 值和 sig.之间是存在一一对应关系的。

问题 12:

怎样判断哪些研究情景适合参数检验，哪些情景适合非参数检验呢？样本量较大(大于 30)时

是不是要用 Z 检验？

回答：

如果数据不能满足参数检验的前提标准，那么就要使用非参数检验。样本量大于 30 的时候也可以使用 t 检验的，样本量越大，t 检验结果越接近 Z 检验结果。

问题 13：

请问选择参数检验还是非参数检验是不是取决于因变量的测量尺度？

回答：

决定采用参数检验还是非参数检验有多个判断标准，变量的测量尺度是其中之一，还有数据的正态性、数据的缺失情况等等，非正态的数据、缺失过多的数据也可能要采用非参数检验。

问题 14：

t 值的计算公式是什么呢？

回答：

t 值的计算公式是：(平均数 1 - 平均数 2) / 标准差。依照不同的情况，标准差的计算方式是不一样的，就不具体展开讲了，因为大多数时候我们不需要进行手算。

问题 15：

在进行 T 检验之前，有必要先检验正态分布吗？

回答：

最好进行正态分布检验。但是如果总体确定是正态的话，t 检验对样本正态性的要求不是那么高。

第 6 讲 比较：它们有何差异？

问题 1:

如果想对使用了不同教学策略的两个班级进行成绩分析，两班均进行了前测后测，这种差异需要做什么数据验证？

回答:

该研究中涉及两种因素，因而建议进行 ANOVA 方差分析。

问题 2:

比如我想研究语言水平对同伴反馈的影响，把学生水平分为高中低三个层次，根据水平差异进行 2 类配对，即同等水平配对 H-H 组，M-M 组，L-L 组和水平不等配对 H-M，H-L, M-L 组，比较两类配对在反馈质量是否存在差异，step2 把同等水平配对的三组进行比较，找出反馈质量最佳的那组，以及水平不等配对同样处理得出反馈质量最佳的一组 step3 把这两组，假设是 H-L 和 M-M, 进行比较，看到底是哪种配对方式的反馈效果最好。这样设计是否合理？

回答:

只要你前期配对好了，数据处理中使用 ANOVA 即可。自变量为两个：配对方式 2（同等水平配对、不同水平配对）X 组别 3（高、中、低）。

问题 3:

根据第六讲的内容，我的理解是 text readability 是 text difficulty 的一个指标，但是 readability 究竟是什么？它需要通过什么来体现 / 测量？

回答:

文本可读性与文本难度，对于这两个概念要从不同角度来理解。文本可读性主要从读者的阅读角度来看，文本难度主要从文本的使用目的来看。因此，两者在一定程度上是相互交叠的。

问题 4:

如果我理解正确（text readability 是 text difficulty 的一个指标），那么 text difficulty 还包括 readability, word frequency, sentence length 以外的哪些指标呢？如果我的理解不正确（text readability 不是 text difficulty 的一个指标），那 text readability 与 text difficulty 是什么关系呢？

回答:

关于文本可读性，给您推荐一篇参考文献，文献中包含了可读性的概念解读和测量计算。阅读后，相信您一定会有所收获。

问题 5:

报告方差分析结果时，自由度 1 和自由度 2 怎么算的呢？在 SPSS 的结果窗口哪里能找到吗？

回答:

在单因素方差分析中，自由度 1 是组数-1（所以如果是两组比较，自由度就是 1；3 组比较，自由度就是 2）；自由度 2 是总样本量-组数。SPSS 结果里会呈现的，是单因素方差分析表中的 df 栏。