

kaobook 类

使用此文档作为模板

示例及说明文档 of the kaobook 类

根据自己需要定制本页

Federico Marotta *

November 9, 2019

奥色姆曼出版社

* A \LaTeX lover

kaobook 类

Disclaimer

你可以编辑这个页面来满足你的需要。例如，这里有一个无版权的声明、一个版权页标记和其他一些信息。这一页是基于肯·阿罗约·奥赫里的论文的相应页面，改动很小。

No copyright

© This book is released into the public domain using the CC0 code. To the extent possible under law, I waive all copyright and related or neighbouring rights to this work.

To view a copy of the CC0 code, visit:

<http://creativecommons.org/publicdomain/zero/1.0/>

Colophon

This document was typeset with the help of KOMA-Script and L^AT_EX using the kaobook class.

The source code of this book is available at:

<https://github.com/fmarotta/kaobook>

(You are welcome to contribute!)

Publisher

First printed in Jan 2019 by 奥色姆曼出版社

世界的和谐体现在形式和数量上，自然哲学的心和灵魂以及一切诗歌都体现在数学美的概念上。

– D'Arcy Wentworth Thompson

前言

我的观点是，每个 \LaTeX 极客，至少在他的一生中有一次，都觉得有必要创建自己的类：这就是发生在我身上的事情，这就是结果，但是，这应该被看作是一个仍在进行中的工作。实际上，这个类并不完全是原创的，但它是我在许多指南、教程、博客和 text.stackexchange.com 文章中发现的所有最佳思想的混合体。特别是，主要的想法来自两个来源：

- ▶ [Ken Arroyo Ohori's Doctoral Thesis](#), which served, with the author's permission, as a backbone for the implementation of this class;
- ▶ The [Tufte-Latex Class](#), which was a model for the style.

本书第一章是导论，涵盖了本课程最基本的特点。接下来，有一堆章节专门讨论所有的命令和环境，你可以用来写一本书；特别地，它将解释如何添加注释、图形和表以及引用。第二部分讨论页面布局和设计，以及其他特性，如彩色框和定理环境。

我开始写这门课是作为一个实验，因此它应该被视为。由于它一直是我个人使用的缩进，它可能不是完美的，但我发现它很满意的用途，我想让它。我分享这篇文章，希望有人能从这里找到写作的灵感。

Federico Marotta

详细目录

前言	vii
详细目录	viii
因素实验设计与方差分析计算原理	1
1 实验设计基本思想	3
1.1 科学研究	3
1.2 心理学实验设计基本问题	9
1.3 实验中变异的控制	15
1.4 假设检验的基本思想	16
1.5 实验中变量的实别	16
2 单因素实验设计	19
2.1 单因素完全随机实验设计	20
单因素完全随机设计的基本特点	20
单因素完全随机实验设计与计算举例	21
2.2 单因素随机区组实验设计	23
2.3 单因素拉丁方实验设计	23
2.4 单因素被试内实验设计	23

插图目录

1.1 Perani(2011)1	5
1.2 Perani(2011)2	5
1.3 Perani(2011)3	6
1.4 Li(2007) 对 12 到 30 个月的汉语母语儿童进行了词汇量的追踪, 结果发现了他们身上也存在词汇爆炸 (vocabulary spurt) 现象	7
1.5 用 LENA 研究父母多言寡言与受教育程度与孩子语言发展情况, 结果发现这二者都和孩子的词汇发展成正相关	8
1.6 荷兰火车的配色, 实际上的车型并非如此	8
1.7 Hagoort(2004)	9
1.8 字的类型 \times 年级	10
1.9 字的类型 \times 熟悉性	10
1.10 能力 \times 熟悉性 \times 字的类型	10
1.11 儿童在各类字读音中犯不同类型错误占总错误的百分比	10
1.12 儿童读音中使作线索的平均次数	11
1.13 Lin(2011) 通过	13
1.14 Boscardin(2008)	14
1.15 Tipper(1991) 发现了基于物体的注意在返回抑制中相比基于空间的注意更具优势	15
2.1 红色是原始数据, 蓝色代表组内误差不变, 加大处理效应; 绿色代表组间变异不变, 改变组内误差	19

表格目录

2.2 保持 SSA 不变, 减小 SSE	20
2.4 保持 SSE 不变, 增大 SSA	20
2.6 单因素完全随机实验设计中被试的分配	20
2.1 原始数据方差分析	20
2.3 绿色数据方差分析	20
2.5 蓝色数据方差分析	20
2.7 单因素完全随机实验的 AS 表	22
2.8 单因素完全随机实验的方差分析表	22

定理目录

因素实验设计与方差分析计算原理

1 实验设计基本思想

1.1 科学研究	3
1.2 心理学实验设计基本问题 . . .	9
1.3 实验中变异的控制	15
1.4 假设检验的基本思想	16
1.5 实验中变量的实别	16

1.1 科学研究

科学研究是揭示物质世界的内在规律，寻求变量间因果关系。牛顿看到苹果自然而然地从树上掉下来，于是他思考，为什么苹果从树上向地上掉而不是向天上掉，当然苹果也可能就停留在空中。这就是物质世界的内在规律，这里例子中就是苹果落到地上和引起苹果掉落的原因。牛顿提出其原因是因为有重力，这个力的作用导致了苹果没有停在空中或者向空中掉的可能，这样把物质世界的规律找了出来，而且这个规律是通过引起与被引起这样的因果关系联系在一起的。同样地，心理学是研究人是如何认识并获取客观世界的知识的，并对人类的行为进行描述、理解、预测和控制。

鲁迅先生曾经写过：“一个人要想离开社会而生存，那正像拔着自己的头发想离开地球一样不可能”。心理学研究也是这样，我们用自己的大脑去研究存在于我们大脑中的心理，这种研究过程本身就是“纠着自己的头发离开地球”，看似我们完全不可能认识自己，但是在这么多年的沉淀中，已经积累了非常多的研究成果，这些进展都在告诉我们，人类的心理是可以研究的，我们是可以逐步深入地研究自己的心理的。下面以语言学习的例子，看看这种人类特有的符号系统人们是如何一步步进行研究的。

关于语言学习，其实每一个人都会有自己的体会，因为每个人都有语言。语言是人类最特殊的一种能力，要了解这种能力，只能通过人来做研究。虽然动物也有简单的交流，但是我们不可能通过动物来研究人的语言。科学家用什么样的方法手段，怎样去研究人类的语言呢？在过去的二三十年中，发生了巨大的变化。最初的时候，我们用行为的方法去研究，收集一些语言的现象，探讨语言的行为，通过反应时手段，推测大脑中语言加工的时间进程。近年来，研究者开始用一些非常精密、非常先进的仪器，包括眼动、计算机模拟、脑电、脑成像、基因的方法，这些方法让我们对人类语言的认识，特别是对语言学习发展、障碍及其脑机制的认识，有了非常重大的突破和进展。

儿童的语言学习是从什么时候开始发生发展的？如果有人接触过孩子的话，可以发现，孩子的语言发展是非常迅速的。从刚出生的时候一点都不会说，到一、二、三岁以后，会说大量的词汇、语句。语言学家、心理学家已经对这个现象进行了大量的探讨，但到现在还是不能完全解释。学前儿童阶段通常是没有正规的语言教学的，但是语言的发展为什么这么迅速？动力在什么地方？机制是什么？这是语言学家、心理学家非常感兴趣的。另外，我们都知道，身高、体重是可以测量的，语言发展可以测量吗？有可能科学地测量哪个孩子语言发展得好，哪个孩子语言发展得差吗？我们是不是可以从儿童在学前时期的发展来预测孩子上学时候的阅读表现？特别是有语言和阅读困难的孩子，我们是否可能早期预测？这些都是特别吸引人的研究问题。

儿童语言是如何习得的是各国心理学家长期感兴趣的问题，其中一个重要的发现是，儿童在 17~18 个月左右有词汇爆发 (vocabulary spurt, Bloom, 1985) 的现象，即从 17 个月之前只能说 50 个词以内，

到 20 个月能说出 500 或更多的词。词汇爆发的机制是什么？按照传统行为主义的观点，人毫无心理可言，所谓心理只是刺激与反应间的联结，所谓学习只是一个简单的刺激-反应见不断加固的过程，语言学习就成了在概念与词语间建立联结，并不断将这个联结强化的过程。但是按照这样的说法，词汇的增长速度应该是平稳增长的，不会出现这样的爆发现象。显然，人类的语言能力显然比行为主义认为的要复杂得多。

Noam Chomsky (1957) 提出大脑里有专门的语言装置，叫 LAD (Language Acquisition Device, 语言获得装置/语言习得装置)，人出生时就具有了掌握语言的天赋，外部语言环境会给语言装置设定参数，以便人类具体掌握一门语言。这个理论很好地解释了“词汇大爆炸现象”，当个体的 LAD 成熟时，就会开始接受外界的语言刺激并且极大丰富个体的语言能力。不过遗憾的是，将近一百年来的研究并没有找到 LAD 在哪里。我不禁想问：这种假设还有可能被证明吗？以前我们认为几乎是不可能的。现在，脑科学发展以后，这个答案被揭开开始变得可行了，我们越来越接近了解更多的东西。比如下面这个实验，揭示了新生儿一出生时他的语言及其脑基础是什么样的，是一片空白？还是说人类先天就能把语言与别的声音区分开来？

Daniela Perani 等 (2011) 给出生刚刚两天的新鲜婴儿做 FMRI，让他们在睡觉的时候听故事，听的是女声讲的母语故事。这个实验设计有三个条件，第一个条件是给新生儿听的是正常的母语故事；第二个条件是新生儿听到的是只有声调，即将正常的声音去除了声母、韵母，只剩下声调；第三个条件是把声调拉平，音节、声母、韵母仍然保持。科学家希望了解当新生儿听这三种不同的声音时，他们的大脑会有反应吗？在哪里反应？

结果在图 1.1 中可以清楚地看到，如果在只有声调的条件下，新生儿大脑两侧都不反应；如果是拉平声调的条件，没有声母、韵母，大脑也不反应；只有在听正常故事的时候，新生儿大脑才有反应。新生儿是真的能听故事吗？研究表明，新生儿不是真的能听故事，他的大脑主要是对人的正常语音进行反应，对语义是不反应的。

进一步的研究结果如图 1.2，还发现新生儿大脑的主要反应区是在大脑双侧的颞叶，是对语音、基本声音反应的脑区，而语义加工的脑区是不激活的。与成年人的语言加工主要在左半球相比，新生儿的语言加工是双侧的。因此我们至少能知道，新生儿出生的时候，大脑皮层已经开始对语音产生反应了。

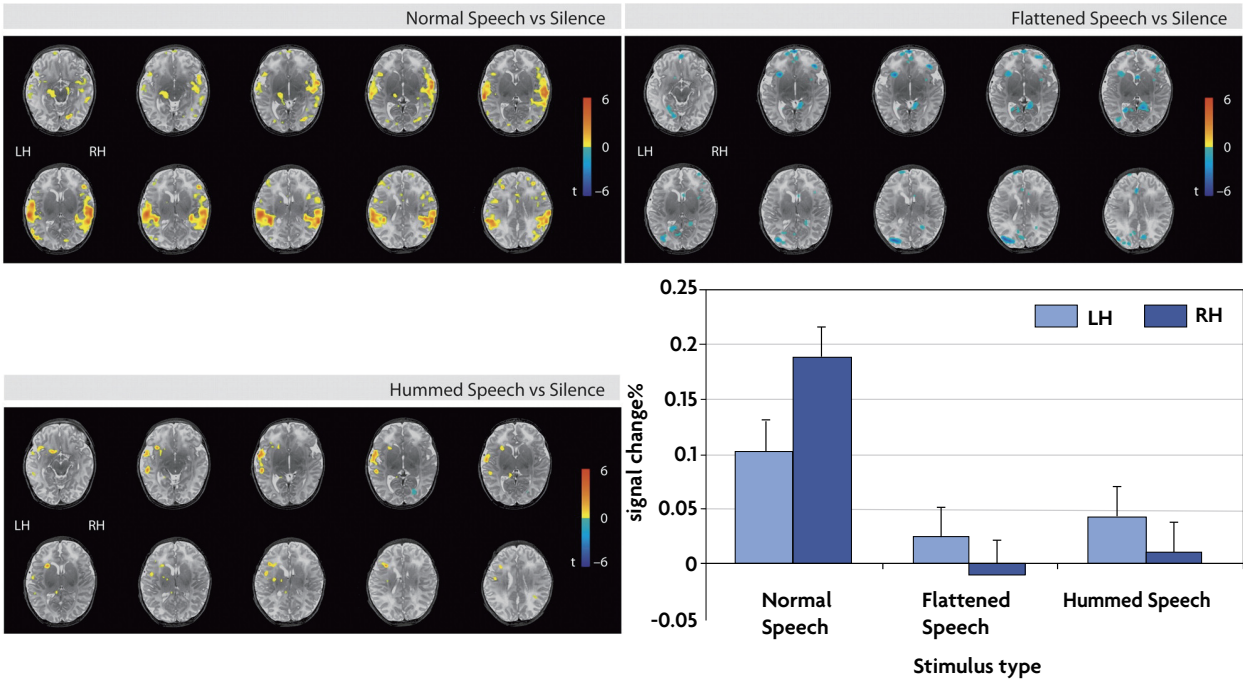


Figure 1.1: Perani(2011) 给两天大的婴儿听三类声音，记录他们的脑部激活情况，结果显示婴儿天生对语言有特异性加工，对非言语声音不产生反应。

更进一步的研究中，Perani 对比了新生儿和成年人听觉中枢的发展情况，如图 1.3中看到的那样，成年人的颞叶和额叶 (A 中绿色和黄色区域) 见有丰富的白质纤维束连接 (蓝色区域)，而新生儿额叶和颞叶间的联结确不那么紧密，这表明婴儿只对语音起反应，几乎没有对更高级的语义和语法加工，而建立起这个联系的主要途径就是接受丰富的语言刺激。

上面这个系列实验是在实验室中通过严格的控制得出的结论，事实上，并不是只有在实验室中的实验才是科学实验，重要的是科学研究的思想。我们先简要说说**科学研究的思想**，随后看看非实验室的研究可以怎样的出科学的结论。

- 1. 系统的理论框架** 只有有了系统的理论支持，收集数据才不是盲目的，上述实验中如果不是有 Chomsky 的理论支持，我们不会新生儿的语言获得情况有所兴趣，如果不是因为有 FMRI 研究技术的发展，我们不会想到用多功能核磁对新生儿进行研究。毕竟，一个刚出生两天的小婴儿，要其主观报告或者用其行为指标来反应其是否对语言有反应是不可能的。研究不是没有目的收集数据，而是有一定的理论支持，有一定的研究目的；
- 2. 控制机制** 主要是说的对无关变量进行控制，这一点和第三条一起说比较方便；
- 3. 对象之间的因果关系** 因果关系是心理学研究中最想得到的，想要做到为点需要有严格的控制。最理想的情况是研究者操控自变量在不同水平，并且控制无关变量，再对因变量进行观测，这样获得因变量观测值的差异就是由研究者在实验中让被试产生的，就可以很肯定的说是由自变量的变化导致的。关于这点我们还要再以后自变量的分类上谈，但此处要留个印象：最严格的因果关系是建立在实验中研究者人为造成的使被试形成的差异，比如上实验中输入的语料类型让婴儿由研究者的操作而导致的它们语言加

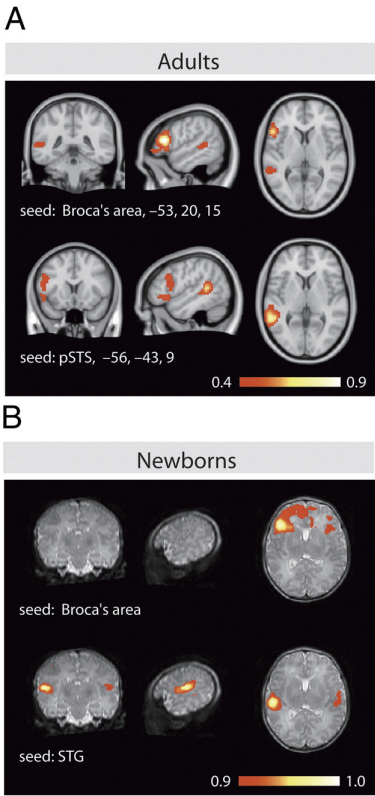


Figure 1.2: Perani(2011) 发现新生儿的语言加工主要是双侧，而成年人的语言加工主要是左半球

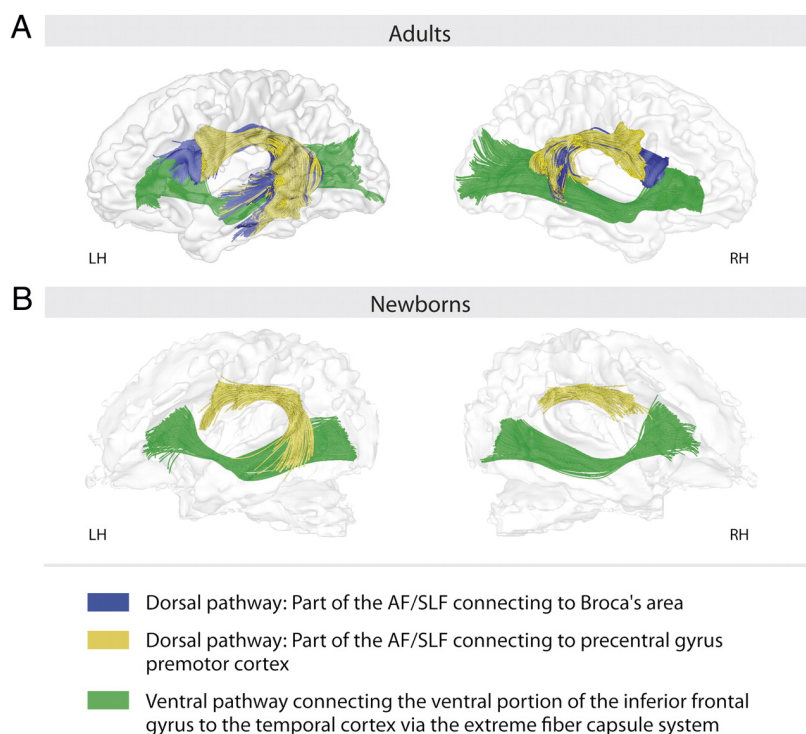


Figure 1.3: Perani(2011) 对比了婴儿和成人听觉中枢的情况，发现婴儿颞叶到额叶见的联系只是初步建立，还不完善，由此推断新生儿只是听声音，但并不明白其含义。

工产生差异，这种差异是在实验室内产生的，是研究者明明白白控制了的，动因明确的。有些被试间的差异不是由研究者引起的，造成这个差异的原因也千奇百怪不得已而知，因此这类自变量往往不能得出严格的因果关系，比如说性别。因为行为学实验个体差异的关系，因此研究者通常用被试内设计，也就是一个被试要接受所有的处理，这样可以有效分离出被试间个体差异，这点我们后面要细谈，此处提及只是因为上实验 (Peranti,2011) 用的就是被试内实验设计，每个婴儿都听了这三段音频。

接下来我们以“对儿童说谎行为的研究”谈谈**科学研究目标**：

- 1. 描述**：如观察儿童说谎行为的各种不同表现形式、说谎的频率；
- 2. 理解**：了解儿童说谎的原因（是否为了逃避批评或惩罚、或者为了获得某种利益？）. 我们知道儿童想要说谎必须有一定的认知机制，说谎不是一个简单的事情，要有一定的认知能力才能说谎。要知道别人对问题是怎么想的，考虑别人怎么看问题才能说出来看起来比较严密的谎言，认知能力还没成熟的孩子说的谎在成年人看起来往往是可笑的；
- 3. 预测**：什么样的家庭或学校环境下，什么年龄儿童可能出现说谎行为；
- 4. 控制**：如果已经知道孩子在青春期喜欢说谎，并且这个现象不利于孩子的发展，那是否可以有合理的控制机制让说谎情况降到最低呢？

下一步，我们继续讨论**心理学研究的途径**，此处不会详细论述某种研究手段，想向各位传达的信息是，不是只有典型的实验研究才是研究，别的研究手段尽管没有对实验条件进行严格的控制，但这些手段都是科学的，都是为了一定的研究目的服务，也会进一步深入问题提供一个基础。额外想提醒的是，研究手段和研究目的紧密相关，有些现象没法用实验的手段，这点我们也放到后面的实验设计中谈。

1. 观察

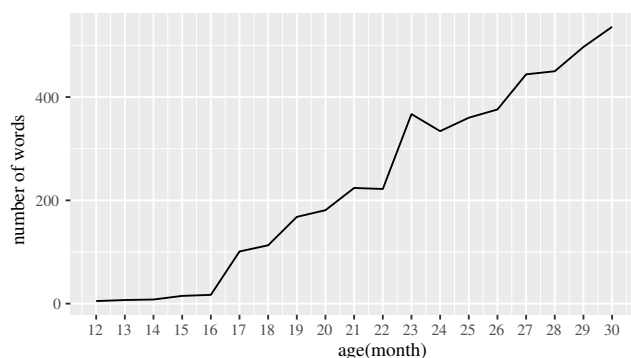


Figure 1.4: Li(2007) 对 12 到 30 个月的汉语母语儿童进行了词汇量的追踪, 结果发现了他们身上也存在词汇爆炸 (vocabulary spurt) 现象

观察法指的是在自然状况下收集数据, 描述某种现象包括自然现象, 包括自然观察, 问卷调查, 个案研究. 前面我们提到了儿童一岁半左右一个很有趣的词汇暴增现象——**词汇爆炸 (vocabulary spurt)**. Bloom(1985) 首先在国外儿童中发现了这个现象, 那么中国儿童是否也存在这个现象呢? 这就有赖于对汉语母语儿童观察, 我国学者确实发现在汉语母语儿童身上也发现了相同的现象 (Li, 2007), 下面介绍一下他们的研究. 我们没法对 6-7 个月的儿童进行控制, 让他们主观报告自己会说的词汇, 因此只能让家长填写词汇问卷, 判断哪些词是自己孩子说过的. 这个词表数目还挺多的, 让一个家长做完整个词表会造成较大的疲劳效应, 因此实验中也匹配了一些“平行家长”, 不过这都是控制机制, 在此我不打算具体介绍. 总之, 在图 1.4 中我们确实看到, 在 18 个月左右时, 确实存在一个词汇量剧增的时间点.

这个实验中, 我们不是被动的收集数据, 而是有着某种特殊目的收集儿童的词汇量, 并且这个实验结果得到了一个漂亮的结果. 可见, 描述性实验也失为一种有用的科学实验

2. 相关研究 相关研究对两种现象 (变量) 之间的关系感兴趣, 例如: 吸烟与肺癌之间的关系. 但是没法指因两个变量间引起与被引起的方向 (到底吸烟引起肺癌, 还是有肺癌基因的人倾向于吸烟不能从显著的相关中推论)

在对儿童语言发展的研究中, 我们感兴趣的是孩子抚养人 (特别是母亲) 的语言使用清空对孩子会有什么样的影响, 具体来说, 父母寡言是否孩子也会寡言? 父母受教育程度高是否导致孩子更喜欢说话? 处于伦理道德的限制, 我们只能记录孩子和父母的情况, 不能对上述两个因素进行严格控制.

这个问题一直不能好好解决, 因为要对一段录音中词语数进行统计是相当困难、工作量巨大的事情. 不过 LENA 技术的出现让这种困境得以摆脱. 这个程序可以识别婴儿的声音中的语言成分, 并把一些非言语部分分离出去, 比如哭声、吃东西的声音; 它还可以分辨不同人的声音, 这样也可以家长的说话情况.

这个实验的结果如图 1.5 所示, 我们能非常清楚地看到, 父母地寡言和多言确实影响了孩子的与发展,

3. 实验研究

以 Hagoort(2004) 的一个研究为例看看他精细的设计. 我们已经知道语义违反¹ 会引起 N400. 还有一类句子错误, 虽然语法、语义上都是正确的, 但是和我们关于这个世界的认识是相违背的, 比如 “中国国家主席是位女士”. 上述的两类错误涉及两种句子理解过程, 一

1: 语义违反 (semantic violation), 指的是合乎语法要求却不符合词语的意义, 如 “这本书很咸”, 从句法来看这句话是正确的, 但是 “咸” 这个形容词不能修饰不是食物的名词, 因此此处属于语义违反句子

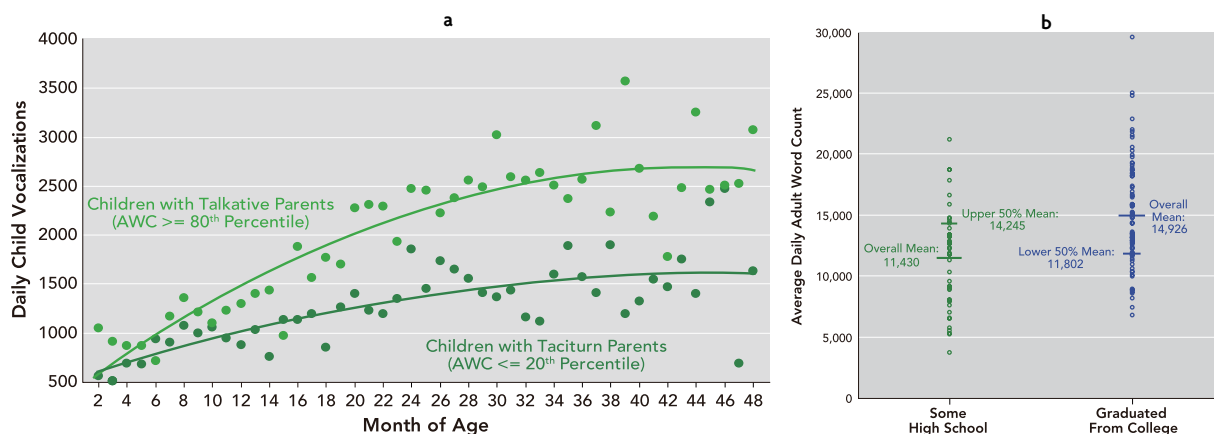


Figure 1.5: 用 LENA 研究父母多言寡言与受教育程度与孩子语言发展情况，结果发现这二者都和孩子的词汇发展成正相关

一个是语义理解，一个是世界知识理解，以往人们认为这是两个独立的过程。Hagoort 指出实际上这两种过程的区别不太好分开，因为很多词语都是不止有一个意思，多义词在特定语境中是什么意思要用世界知识来判断，进而 Hagoort 判断，世界知识违反 (world knowledge violation) 也会引起 N400。

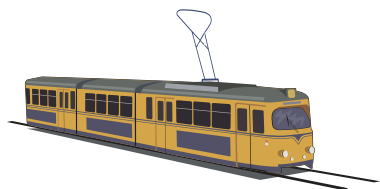


Figure 1.6: 荷兰火车的配色，实际上的车型并非如此

实验的材料 (见图 1.7 中的文字) 经过了精心的设计，Hagoort 给世界知识违反、语义违反、正确句子下了三个操作定义，这里用了一个非常有意思的操作定义来表示世界知识违反：实验的被试都是荷兰人，荷兰的火车都是黄色为主色调的 (配色类似图 1.6)，因此 “The dutch trans are white” 显然不符合荷兰人心中存在的对火车的知识。结果非常证实了 Hagoort 的猜想，在图 1.7 可以清楚地看到，正确句子没有引起 N400，而语义违反和世界知识违反都引起了 N400。

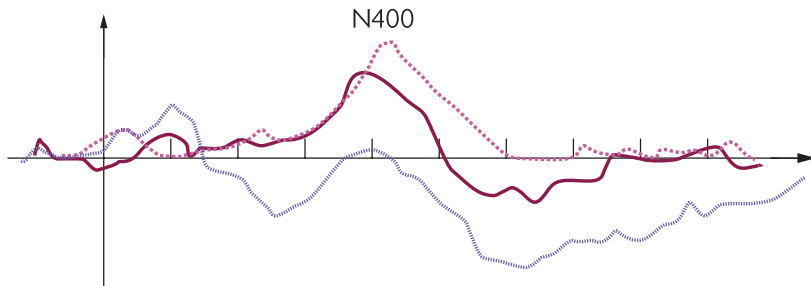
问题：如果这个实验只保留世界知识违反和正常条件，可行吗？

如果顺着这个实验的结果来推理，这么做是看起来是可行的：把这个实验结果上的语义违反条件去掉，世界知识违反也确实引起了 N400。但是要注意，实验前研究者并不知道结果是什么。也就是说实验前的假设是：

- (1) 若世界知识违反可以引发 N400，说明语义理解与世界知识理解是一个连续的过程；
- (2) 若世界知识违反没有引发 N400，说明语义理解与世界知识违反是两个独立的过程。

如果没有语义违反条件，(1) 的结论还是可以得出，但是 (2) 中似乎存在了点问题。因为实验结果不显著不一定是处理效应不存在，还可能是检验力低的问题。对于心理学实验，它的处理效应本身就不高，那些不太明显但是确实存在的现象才有研究和发掘的价值。因此处理效应的显著需要较高检验力，实验设计一定要尽力精准和敏感。但不是所有实验都可以做到那么精准，想要得到漂亮的结果不是一件容易的事。在本实验中，如果没有语义违反做对照，在世界知识违反不显著时，研究者没法区分这是由实验不敏感导致的，还是由处理效应不存在导致的。因为语义违反已经被无数实验证明一定会引起 N400，所以加上语义违反这个条件，就可以对实验的敏感性进行检验。这样 (2) 就可以拆分为以下两个假设：

- (1) 若世界知识违反没有引起 N400，语义违反没有引起 N400，则



The Dutch trains are yellow and very crowded.
 The Dutch trains are white and very crowded.
 The Dutch trains are sour and very crowded.

Figure 1.7: Hagoort(2004) 给被试看三个句子并记录其脑电，发现世界知识违反的句子也会引起 N400，进而说明了语义理解与世界知识理解是一个整体而不可以分割的过程。

说明实验不够敏感，需要重新设计实验，处理效应是否存在有待继续检验；

(2) 若世界知识违反没有引起 N400，语义违反引起 N400，则说明世界知识范围确实不会引起 N400，世界知识违反和语义违反是两个独立的过程。

分析一下这个实验为什么是一个好实验。首先在自变量研究者控制了关键词的不同，在无关变量的控制上，除了关键词以外的句子都一摸一样，这样就排除了由于文字本身难度、熟悉度、使用频率带来的无关变异。实验设计上，采用被试内实验设计，每个被试都读了这三个句子，控制了所有的被试间的差异。这样总得来看，操作定义下得很精准、无关变量控制得也很巧妙，是一个值得我们学习的非常典型的实验设计。

1.2 心理学实验设计基本问题

从上面的实验可以看到，**心理学实验的特点**是：

(1) 影响因素多；(2) 因素之间相互关联；(3) 无关变量的混淆

研究者在操纵和改变自变量时候，有大量无关变量，有这么多复杂的因素需要控制，行为实验中最复杂的因素，这也是以人研究对象的实验的难点。当然我们也非常自豪地说，这是心理学家最擅长的。

我们接下来讨论什么是**实验设计**。广义的实验设计涉及的内容比较广泛，此处不予以介绍，我们看看狭义的、典型意义的实验设计要做什么。狭义的实验设计就是做好实验的计划方案、统计分析方案，就像盖房子需要一个蓝图，我们不能在盖房子前什么规划都不做就直接去盖，同样我们也不能没有计划就去进行一个实验。实验之前要做一个详细的规划，这非常重要。如果这个方案没有做好，之后便很难纠正。

实验设计的主要目标有三点，首先，要回答感兴趣的理论问题；其次，要获得更加丰富的信息；最后，要提高的敏感性。我们分别来看看这些目标的具体含义：

1. 回答感兴趣的理论问题 前面已经提到了，我们感兴趣的问题是**人类心理现象的规律性问题**，也是对人类行为和心理现象产生的**动因**感兴趣。所以首先我们要建立变量之间的关系，在实验中操纵或改变一个或多个变量，并观察其对行为的影响，记录下因变量的观测指标的变化情况。

但是影响某个心理或行为的因素不只有我们感兴趣的自变量那么简单, 我们的想法是让不同被试仅仅在自变量上有所不同, 并且这个不同是由研究者操纵的, 这样结果就确定是由自变量引起的. 但由于实验对象的特殊性, 我们很难保证不同人在进行实验前是完全相等的, 所以我们还要对那些对因变量也产生影响, 但是不是为研究者所感兴趣的变量进行控制, 也就是要控制无关变量, 减小或控制其对因变量影响.

最后, 选用复杂的统计方法, 在同样的数据量中得到更多有意义的信息.

2. 更加丰富的信息 首先, 可以选用更好的实验设计的种类, 比如在多因素设计中, 除了主效应, 还有交互作用, 而交互作用往往可以提供一些非常有用的信息. 在研究中, 不同因素间往往存在相互影响, 忽略一个因素而研究另一个因素时, 难以看到真实的结果, 这就是单因素实验设计的局限.

其次, 可以选用丰富的因变量, 这个等下举个例子说明.

最后, 正确选择一些复杂的统计

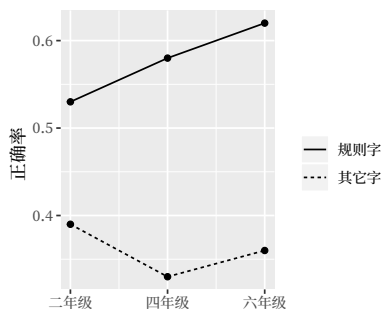


Figure 1.8: 字的类型 × 年级

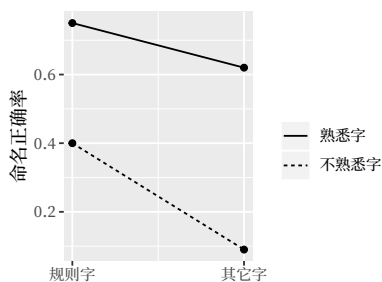


Figure 1.9: 字的类型 × 熟悉性

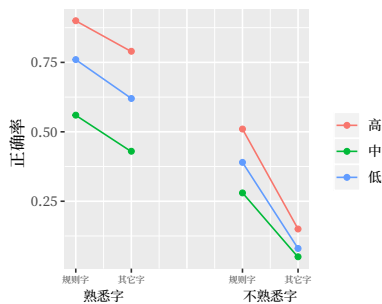


Figure 1.10: 能力 × 熟悉性 × 字的类型

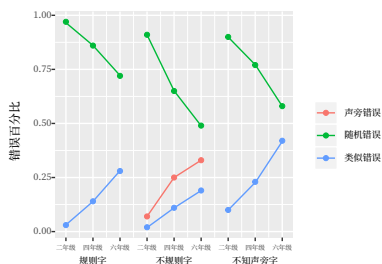


Figure 1.11: 儿童在各类字读音中犯不同类型错误占总错误的百分比

多因素实验设计的优点

(1) 可以同时估计两个或多个实验处理的效应, 可以估价实验处理之间的交互作用

(2) 因素实验设计可以对数据资源进行更加有效的利用, 使同样数量的数据给出丰富的信息

但是要注意的是, 多因素实验设计的统计和解释可能会比较复杂, 特别是对于一个高次的交互作用, 很难解释其意义; 并且, 过多因素的实验设计可能设计时觉得还挺漂亮的, 但是实际做出来交互作用都不显著. 这提醒我们研究一个问题的时候要先深入研究该领域, 当有足够理论支持多因素实验的交互作用显著时, 多因素实验设计才可能成功.

3. 提高实验的敏感性 前面我们已经提到, 处理不显著的时候, 可能真的是处理效应不存在, 也可能是实验设计不够敏感, 我们要尽量减小后一种的可能性. 不过大多数时候如果处理效应不显著, 我们并不能将这两者区分开来.

舒华 (1996) 的一项研究中, 对孩子读汉字的过程进行研究, 特别是不熟悉的汉字. 研究者想看到孩子读汉字能力的发展. 实验中有一个非常漂亮的三次交互作用, 揭示了一下有价值的信息.

这是一个 3(年级)×3(语文能力)×2(字的熟悉性)×3(字的类型) 的四因素混合设计, 其中字的熟悉性和字的类型是被试内变量, 语文能力和年级是被试间变量. 实验材料是 60 个单个形声字组成的卷子, 所有的字都选自小学语文课本, 每个年级的卷中包括 30 个熟悉的字和 30 个不熟悉的字. 熟悉字指学生在课本中已经学过的字, 不熟悉字指学生在课本中还未学过的字, 例如四年级卷子中的熟悉字选自第五、六册语文课本 (三年级用书) 后面的生字表, 不熟悉字选自第九、十册语文课本 (五年级用书) 后面的生字表. 每种熟悉度的字中有 10 个规则字, 即字的声旁本身是一个儿童熟悉的汉字, 且声旁的读音与字的读音相同, 如“绘”; 10 个不规则字, 即字的声旁是一个儿童熟悉的汉字, 但声旁的读音与字的读音不相同, 如“略”; 和 10 个不知声旁的字, 即字的声旁不是一个独立的汉字, 其读音是一般人所不熟悉的, 如“忱”, 它的声旁“尤”的读音是“yin2”. 不同熟悉度和不同类型的字的笔划数是经过匹配的, 平均笔划数为 10. 实验是在三个自然班中集体进行

的, 每个学生接受相应年级的卷子, 并按要求用拼音给出 30 个熟悉字和 30 个不熟悉字的读音, 对写不出准确拼音的字, 可以写一个与它同音的熟字代替。

上述几个自变量比较难操作的是熟悉性。熟悉是最熟悉、全部都认识的吗? 不熟悉、一定不认识的吗? 这样做的话会出现天花板和地板效应, 也就是熟悉的字成绩为 100%, 不熟悉的字成绩为 0。比较好的实验设计是, 熟悉是比较熟悉, 但也不是百分之百熟悉; 不熟悉是相对不熟悉, 也不是 0。额外, 上面提到, 不同年级的材料间有一定的重叠这样使用的汉字数量比较少, 这就减少了由汉字本身特点带来的无关变异。最后二四六年级的儿童匹配到的熟悉和不熟悉汉字的来源是:

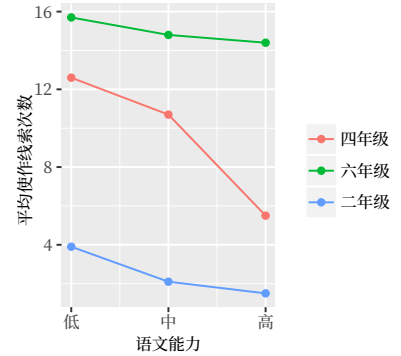
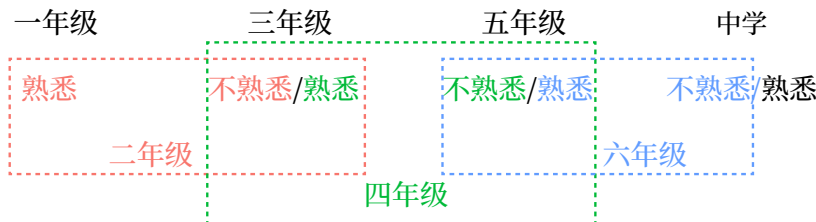


Figure 1.12: 儿童读音中使用线索的平均次数

年级和字的类型的交互作用显著 (见图 ??, 规则字随着年龄增长, 不规则字差不多, 使用这个规则有一个学习过程

字的类型和熟悉性的交互作用显著, 不规则的字更倾向于推理

语文能力三次交互作用显著, 不熟悉的字中

对错误的分析中, 三类随机错误三类错误都下降不规则字声旁错误增加类似错误也是增加的儿童对汉字的认知在发展, 光读声旁很可能读错, 所以在家族中找一个字

使用线索的次数, 四年级有一个较大的发展

正确率由前侧, 50% 正确率有较大变异

想要达到实验设计的目的主要通过以下三种手段:

- 1. 加大处理效应** 有的时候增大处理效应不是一个很好的方法, 因为处理效应太大了会没有意义。比如我想研究人的数学能力的发展, 我选取 3 岁、15 岁、60 岁的人, 可是我在实验前就已经知道 60 岁的人数学能力大于 15 岁的人大于 3 岁的人, 这就没有研究的必要了。心理学的效应往往是日常生活中本身不太明显, 以至于人们一般看不到这个现象, 我们的研究就是要发现这个微弱的效应, 并说明其探讨其存在意义。
- 2. 减小误差** 以方差分析为例, 方差分析最后判断是否显著构成了一个 F 分布:

$$F = \frac{MS_{effect}}{MS_{error}}$$

实验的目的就是想让 F 达到显著水平, 也就是 F 越大越好, 上面已经否定让分子无限变大的可能性, 那么分母是不是可以尽量控制让其减小呢? 答案是肯定的, 控制实验误差使其很小是实验设计的中心问题。

通过实验设计减小误差的方法

- (1) 区组、拉丁方实验设计
- (2) 混合、被室内实验设计
- (3) 嵌套、协方差实验设计

3. **增加被试** 简单要地说, 在心理实验中, 不管效应值多小, 只要增加足够的被试量, 总能达到接受备择假设的结果. 增加样本容量意味着成本的, 在有些心理学实验中这很困难, 因此要注重别的提高实验敏感性的方法.

统计在实验中的重要性

汉语阅读障碍儿童的认知缺陷 Age reading Syslexia control 什么东西导致了阅读障碍?

语素意识、语音意识、命名速度口语词汇 (考察基本认知能力) ——描述性阅读障碍的孩子最基本的认知能力有关

相关

但是不能得到因果 1) 追踪 2) 训练

我们再来看两个研究, 第一个研究室 Eron(1972) 非常著名的研究美国儿童看暴力电视和攻击性见关系的研究, 一个是 Lin(2011) 对汉语阅读障碍儿童的追踪研究.

基于相关可以得到因果吗? 小学生对暴力电视的偏爱和攻击性 Eron(1960)

之前的研究的表明, 学校中表现出阅读障碍的孩子主要是三个基本认知能力不太好, 分别是语音意识、语素意识和快速命名. 这说明阅读障碍的孩子不是由于智力低, 问题出在某些生理层面. 孩子进入小学后再发现他们的阅读能力有障碍, 会导致他们所有学科都不能好好发展, 对于学龄儿童的干预也不如学前那么有效, 因此研究者们想找到在学前可以预测上学后阅读障碍的指标, 这样在学前对他们进行人为干预, 哪种认知能力差就进行专门训练.

Lin(2011) 发表一个 6 年的纵向研究, 追踪了 261 名汉语孩子的基本认知能力的发展, 用混合增长模型将他们分成四个亚类型:

- ▶ **typical**: 阅读能力正常发展的孩子
- ▶ **catch-up**: 一开始基本认知能力比较差, 但而后显示出正常的阅读能力的孩子
- ▶ **literacy-related-cognitive-delay**: 在语音意识、语速意识、快速命名上表现较差的孩子, 这些基本认知能力也导致他们的汉字识别量很低
- ▶ **language-delay**: 所有的认知任务中都发展的相对较慢的孩子

结果 (见图 1.13) 可以看到, 正常发展的孩子 (紫色) 各项认知能力发展都很好; 后来居上的孩子 (红色) 有一些能力在学前发展不太好, 但是上学后有较好的发展; 剩下的两组孩子 (蓝色和绿色) 的总体发展都不太好. 额外, 在 8 岁半的进行的文字识别和阅读流畅度测验中, 后来居上的孩子² 表现也不是特别好. 这样研究者就得到了这样几个可以预测阅读障碍的指标——语音意识、语素意识、快速命名.

2: 研究者发现这些后来居上的孩子其实是因为在学前抚养者不爱说话, 导致孩子接受了较少的语言刺激, 尽管后来他们的各项基本认知能力有所恢复, 但是在后面的两个和阅读和文字的任务中仍然表现的不够好

上面的这个实验提到了一种叫混合增长模型的统计技术, 我们简要介绍一下它. 图 1.14 是 Boscardin(2008) 的一项研究的结果, 这个实验的数据分析就是用的这个模型. 最大的图中, 可以看到这个数学模型

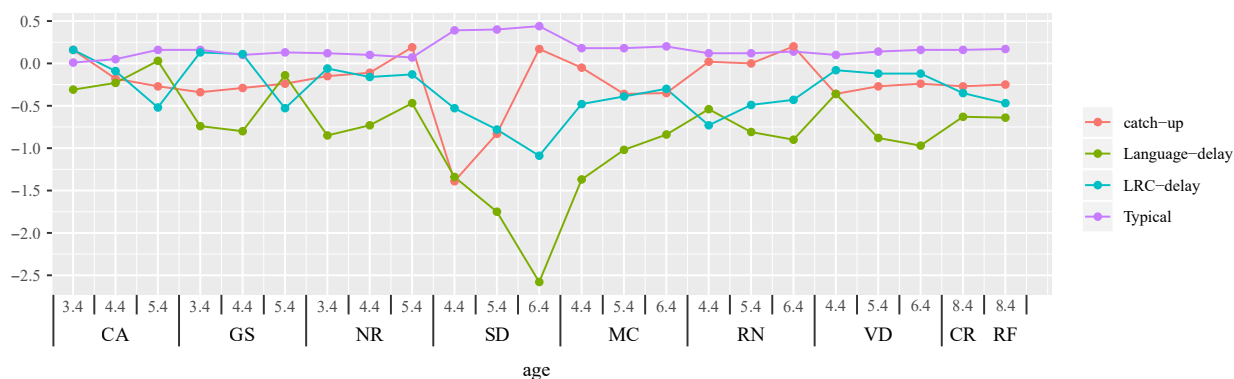


Figure 1.13: Lin(2011) 通过

很好地把孩子分成两个亚类型，类型内的孩子比较接近，但是类型间的孩子相差较远；而且这些孩子中，有些孩子斜率相同，起点不同；有些孩子起点相同，斜率不同；也有可能起点不一样斜率也不一样。这就是用到复杂的统计，得到了非常丰富的信息。所以对于数学和统计学习一定要下尽全力，我们不仅要有最精细的实验设计，也要有最复杂的统计分析手段，这样才能做最好的研究，只是大多研究者醒悟的比较晚，若你现在还有心去学数学，那就立马去学吧！

前面介绍的部分实验用到了比较新的技术，一方面在工具上，针对于脑研究，像 ERP/FMRI/眼动仪都是非常好的研究工具，它们帮助我们更深层次探讨人脑的奥秘；另一方面，新的统计技术也让研究可以更精确的得出结果，包括已经写入大多课本的多重回归、路径分析、结构方程，也有上面提到的混合增长模型。不论这些手段多么的眼花缭乱，一定要注意，实质仍然是实验设计。比如下面这个实验，虽然仍然是行为学实验，不过实验设计非常有趣，非常漂亮地得到了结论。

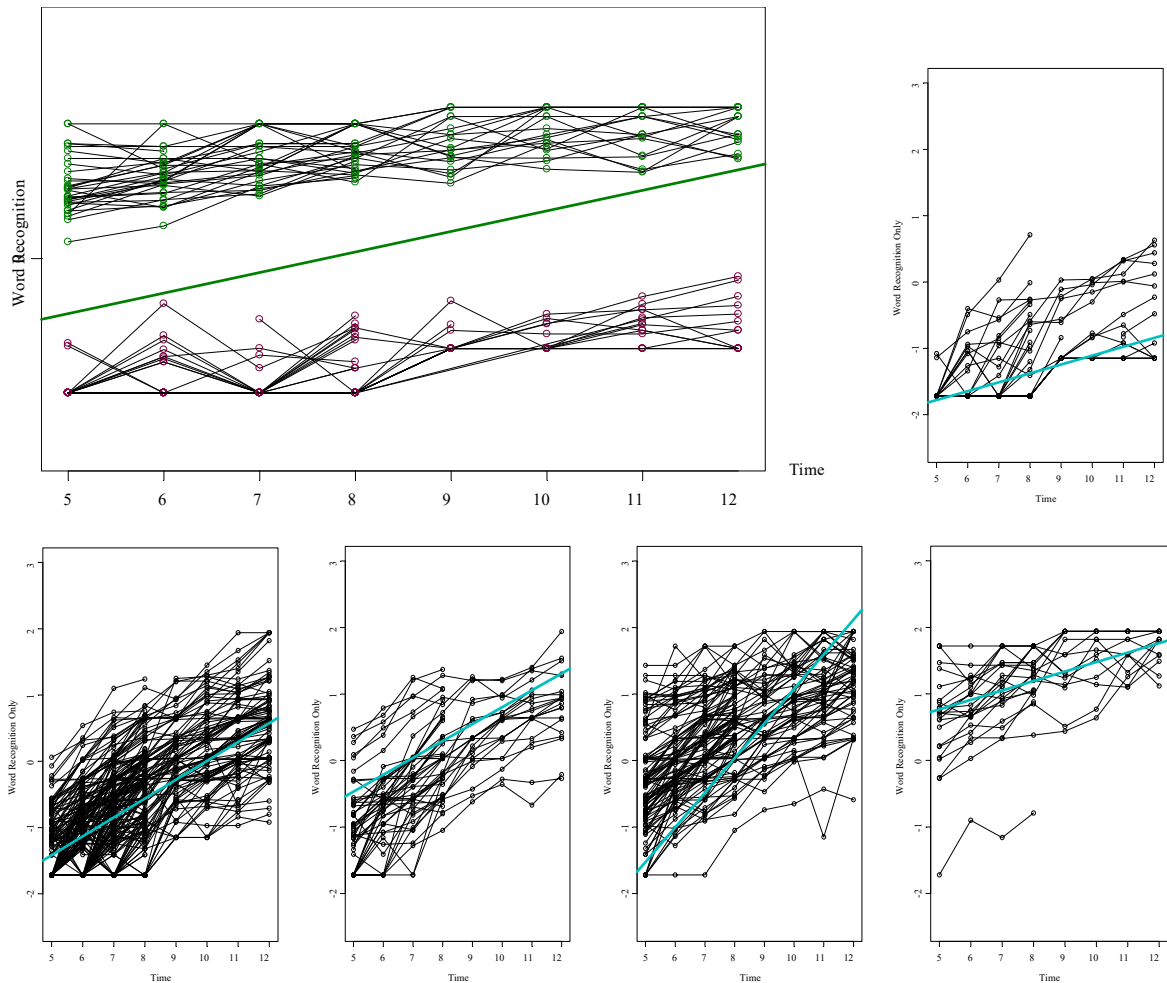


Figure 1.14: Boscardin(2008)

Tipper(1991) 研究了注意中的返回抑制现象 (inhibition of return), 所谓返回抑制, 大体上是说对于我们注意过空间的某个区域, 再次对其注意需要付出更多努力, 在行为上的表现就是需要花费更多时间去重新注意. 这点在学生看书时表现的很常见, 在看完一本书后, 我们好像有一种不知道哪里来的抗拒心理让我们不愿意复习.

Tipper 想知道的是, 我们到底是不愿意再注意注意过的物体, 还是不愿意注意注意过的空间位置? 若二者都存在抑制效应, 哪种注意更占优势呢? 图 1.15是他的研究. 初始状态 a, 有三个方块在视野中, 方块 1、方块 2 和中间的方块. 1 和 2 顺时针旋转到水平时 (b), 1 闪烁, 继续旋转到 c 后中间方块闪烁, 然后等到旋转到 e, 此时 1 和 2 的位置和 b 中正好相反, 然后 1 再次闪烁或者 2 闪烁, 记录这次闪烁到按键间的反应时, 发现被试更加难以注意 1, 说明了基于物体的注意在返回抑制中更有优势.

心理学研究的发展: 实验心理学: 拥有严密的实验控制教育与社会心理学: 大样本 + 统计

复杂实验设计和多元统计的结合

推论统计——从样本推断到总体, 不用 1 有限的现象: 可以观察到全体样本 2 可以直接观察: DNA 螺旋

统计假说：一种推论形式，可基于不完整的信息检验科学假说的真伪

方差分析的基本思想对一组数据的描述：集中、离散 X1: 7 5 4 10 4 X2: 10 1 1 15 3 mean: 6 6 range: 4-10 1-15 var: 26 156

$$t = \frac{X_1 - X_2}{S_P}$$

变异是否存在？

施加处理：一致性变化（组内变化没变，组间有变）

随机分配被试来保证实验前被试同质

$$F = \frac{MS_{effect}}{MS_{error}}$$

下面应该是一个随机误差，所有处理效应要和误差效应相比，

1.3 实验中变异的控制

所谓变异 (variance)，就是对数据离散情况的衡量。和方差 (variance) 是一个词，不过实验中的变异含有三种变异

1. 使系统变异最大

- ▶ 选取适当的自变量水平，使自变量水平的改变所引起的变异能在因变量中反映出来
- ▶ 选择对自变量的变化敏感的因变量

这个实验中，一个是 10,000 条，一个是 20,000 条因变量：容易形成正态分布愿意买哪本？1010 的数据，变化不敏感，变异小 **2. 控制无关变异** 所有可能做自变量的因素均可能成为因素均可能成为无关变异的来源，无关变量是心理学、教育学、社会学研究中最棘手的问题之一。

上面大：例如研究两种教学条件对成绩的影响，实验选择两个班，但是可能两个班的学生能力不同，使得处理效应变异虚大下面大：主要是被试带来的个体差异

这个实验中随机分配指导语的描述一模一样，由不同文字引起的带走了

有五种控制无法变异的基本方法：

- ▶ 随机化：随机分配被试
- ▶ 消除：不是一个好方法，比如性别会影响，就只研究男性
- ▶ 匹配：基本上都是一样的
- ▶ 附加自变量
- ▶ 统计控制

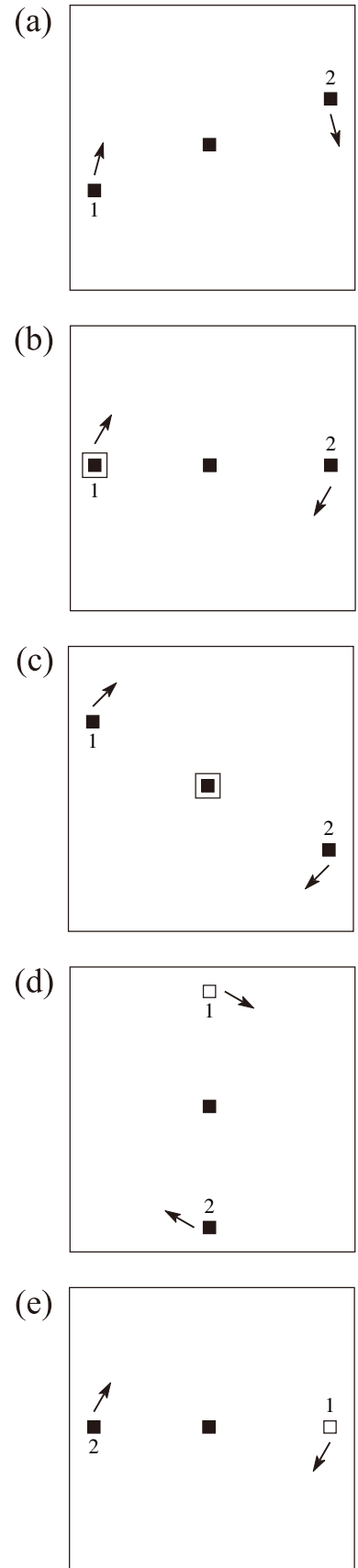


Figure 1.15: Tipper(1991) 发现了基于物体的注意在返回抑制中相比基于空间的注意更具优势

3. 使误差变异最小 误差指的是在实验中没有控制的变异：主要来自于两个：处理单元内被试间的个体差异，我们知道这个是行为学实验中最主要的差异，测量误差尽量在实验设计中控制，实在控制不了的用统计控制

实验设计的评价：内部效度（稳定性、控制无法变异）、外部效度

1.4 假设检验的基本思想

F 分布的基本假设

- ▶ 因变量观测值大体服从正态分布，如果数据不可能符合正态分布，就不能用 F 检验
- ▶ 变异的同质性：由随机分配保证，是一个最基本的假设，因为我们根本就不做前测，只做后测加实验处理只是变了组间变异，实验后的组间变异也反应了处理前的组内变异
- ▶ 独立性：每一个观测值独立于其它观测值

实验设计模型及其假设：变异的构成一次观测就只有一个观察值，但是怎么知道这个观察值有哪些组成的？这个模型就是指导这件事，如：

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{i(j)}$$

$$SS_{total} = SS_{within_group} + SS_{between_group}$$

误差是一个平均数为 0 的随机变量，不是一个定向的变化

实验设计的基本过程：1 提出问题和假说的形成问题特征：探索两个或多个变量间的关系问题的可检验性：研究该问题的方法和手段已经解决或初步解决 Science 290, 1582-1585(2000) 有意思的问题，在以前是没法做的，婴儿没法报告问题所包含的变量是可以检验并可以操作化的（操作定义）怎么操作世界知识违反操作化？非常精巧化

1.5 实验中变量的实别

变量的选择能否回答想解决的问题？符合：1 研究的理论假说/实验设计和统计的要求

被试固有特性：被试在实验前就有的，不太可能改变和操纵，通常分类严格地来说，追踪的研究能得到严格的因果关系吗？最严格的因果关系是操纵和改变自变量，引起心理或行为上的变化有些情况下没有办法做这个但是又没有办法做，已经是最好的了通常在实验中还会加一个可以操控的，不会让自变量只有被试固有特性，

被试的暂时特性：心理学严重最丰富，最具特色的自变量。操纵外部刺激引起心理变化，不是完全外部，引起的心理变化正好是想研究的问题，但是比较难以操纵，需要找到非常好的操作定义。

可操作、可改变的不可操作、不可改变的

自变量的数量：与问题、统计方法有关水平的数量：动机的那个例子

主要意义次要次要意义, 定 SOA, 43 主要意义一直维持激活, 次要意义有一个上升下降的曲线

因变量: 可以看到的必须随着自变量的变化而变化可以转化为数据

1) 对自变量的变化最敏感: 字典中 2) 实验中选取的观测量应该可靠, 得出稳定的结果 3) 大体上正态分布, 人的行为、生理大体上都符合正态分布, 但是如果数据就不可能服从正态分布就不要用方差分析 4) 可行性上考虑

Keenan(2001) 自我面孔实别

无关变量

操作定义: 变量必须用测量或操作它们的步骤来定义用事物可观察、可测量的特征把研究变量具体化, 它清楚地讲明与某些现象独特联系的可观测的准则

意义: 1、对研究结果评价: 同意研究结果必须先接受操作定义 2、有利于研究间交流: 在前人基础上重复或修改实验 3、保障可重复性

获得方式: 1、使用现有信息: 性别、教育程度、智力 2、操作创建一种形态: 让被试进入实验后让其脑子发生变化, 最重要 3、通过前测、评定、记录获得信息, 如语义相关

Rosler(2001)

Loftus(1974) 看相同的视频, 只是问题的词不同 The possible speed when the two cars

心理学研究中的道德问题——诚信 2005 年有一个韩国科学家黄禹锡胚胎干细胞提取编造数据 Science 撤销文章

实验无法重复的原因: 1) 报告不真实: 1990 然后, 美国艾奥瓦大学医学部内科一名高级研究员进行了一系列实验来验证某种物质对毛细血管扩张的影响. 在最后报告数据的时候, 他仅保留了那些支持其假说的数据, 而排除了与假说不相吻合的数据.

美化学家撤销 7 篇论文碳-氢活化实验结果无法重复, 美国纽约哥伦比亚大学化学教授 Dalibor Sames 宣布撤销在《美国化学学会期刊》《有机快报》等杂志上的 7 篇论文. Sames 教授曾被认为是碳-氢活化领域的先锋, 他的研究小组大大推进了该领域的发展. 撤销论文的原因是, 实验室的其他研究人员无法重复文中的结果.

2) 实验控制无关变量掺在 F 的上面还是下面都不好

方法的描述要细致到别人可以重复的地步

2 单因素实验设计

以往在论述实验设计时，我们只交代了做一个随机区组或者用被试内可以带走无关变异这样一个事实，但这只是描述性的话，没有数据的前后对比无法让人真的信服随机区组和被试内设计的好处。所以，现在要从数据上看到 F 怎样更容易显著。

所谓 F 更容易显著无非就是 F 更大了 ($F = MSA/MSE$)，想让 F 变大有两个手段，让处理效应 (MSA) 变大，或者让误差效应 (MSE) 变小，这件事情在前面已经介绍过。我说过心理学家感兴趣的那些现象往往都不太大¹，而且人为加大处理效应有时没有什么意义。从实验设计的角度上，减小误差才是关键。

这就好比用天文望远镜观测外太空，我想发现人们不曾发现过的行星，我透过天文望远镜可以看到一些噪音——没有规律发亮的光点，但是那颗星星好像也在发光，我想鉴别这个星星是否真实存在，但是这件事情不是那么容易，因为星星发的光和噪音总是同时出现，故只有星星发出的光足够大，大到和周围的噪音是那么的与众不同，我再将这个星星是噪音有很大几率犯错，我才敢认定那是星星。我们清楚地知道，想要更加清楚地观察到星星无非有两种手段，其一是让星星在原有基础上多发电光；其二是让周围的噪音降低点。

方差分析模型中的 MSA 和 MSE 是相互独立的，它们俩可以独立地变化。随机误差指的是由随机因素导致数据在理想值周围随机晃动的偏差，它要是和处理不独立那就不能叫随机误差了。

思考

假定我们有一组分数，分别计算出 MSA 和 MSE 。

(1) 我们可以改变 AS^2 分数，保持处理平均数恒定。即保持 MSA 不变而改变 MSE 。

(2) 如果我们改变组平均数，处理组内分数关系不就， MSA 会变化，而 MSE 不就。表明两个均方是独立变化的。

我们看一组数据

a_1 : 1, 2, 3, 4, 5

a_2 : 5, 6, 7, 8, 9

(1) 调整原始数据，保持 SSA 不变，而 SSE 改变 (减小 SSE) (通常加大处理效应实现)

(2) 调整原始数据，改变 SSA 不变 (增大 SSA)，而 SSE 不变

我直接给出我的思路，图 2.1 中红色是原始数据，它有两个水平， a_1 的各个值围绕其均值 3 波动， a_2 的各个值围绕其均值 7 波动。

绿色是我写的符合 (1) 的一组数据，可以看到红色数据 a_1 处理的每个数据都围绕着 3 波动，现在绿色的值在 a_1 内还是围绕 3 波动，但是数据间的距离更短了，它们显得更加紧密。虽然从组内均值上来看

2.1 单因素完全随机实验设计 . . .	20
单因素完全随机设计的基本特点	20
单因素完全随机实验设计与计算举例	21
2.2 单因素随机区组实验设计 . . .	23
2.3 单因素拉丁方实验设计	23
2.4 单因素被试内实验设计	23

1: 也可以说是效应量 (effect size) 总是不太大

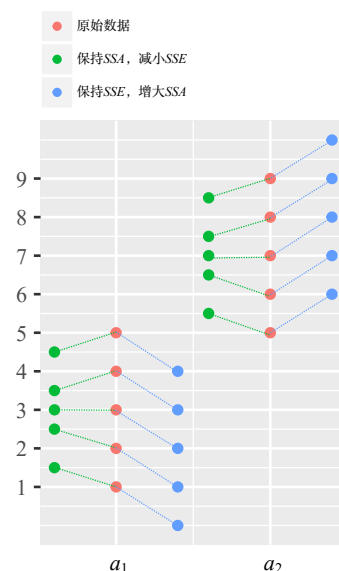


Figure 2.1: 红色是原始数据，蓝色代表组内误差不变，加大处理效应；绿色代表组内变异不变，改变组内误差

2: AS 中 A 代表因素 A 的处理效应， S 代表被试个体误差，具体是什么后面再说

(3 和 7) 两组数据的平均值之差不变, 但是由于 a_1 的数据更集中于 3, a_2 的数据更集中于 7, 故实际上这样两组间的差异其实增大了。

蓝色是我写的符合 (2) 的一组数据, 可以看到数据间的紧密程度不变, a_1 整体向下移动, a_2 整体向上移动, 这样组间的差距被拉大。

原始数据方差分析结果是。

Table 2.1: 原始数据方差分析

变异来源	SS	df	MS	F	p
组间变异	40.000	1.000	40.000	16.000	0.004
组内变异	20.000	8.000	2.500		

Table 2.2: 保持 SSA 不变, 减小 SSE

a_1	1.5	2.5	3	3.5	4.5
a_2	5.5	6.5	7	7.5	8.5

绿色数据见表 2.2。

其方差分析结果是, 可以看到相比表 2.1 中的数据, 处理效应没变, 误差效应减小了, F 值变大, p 值减小。

Table 2.3: 绿色数据方差分析

变异来源	SS	df	MS	F	p
组间变异	40.000	1.000	40.000	32.000	< .001
组内变异	10.000	8.000	1.250		

Table 2.4: 保持 SSE 不变, 增大 SSA

a_1	1	1	2	3	4
a_2	6	7	8	8	10

蓝色数据见表 2.4。

其方差分析结果是, 可以看到相比表 2.1 中的数据, 处理效应变大, 组内误差减没变, F 值变大, p 值减小。

Table 2.5: 蓝色数据方差分析

变异来源	SS	df	MS	F	p
组间变异	90.000	1.000	90.000	36.000	< .001
组内变异	20.000	8.000	2.500		

2.1 单因素完全随机实验设计

单因素完全随机设计的基本特点

3: 处理的组合是多因素实验中会遇到的, 后面遇到了再说

Table 2.6: 单因素完全随机实验设计中被试的分配

a_1	a_2	a_3	a_4
S_1	S_2	S_3	S_4
S_5	S_6	S_7	S_8
S_9	S_{10}	S_{11}	S_{12}
S_{13}	S_{14}	S_{15}	S_{16}

完全随机设计的假定是, 由于被试是从总体中随机选取, 处理也是随机分配给被试, 那么被试间的差异也是随机、在统计上无差异的。

单因素完全随机设计中分配被试的图解例子如表 2.6, 个因素 A 有 4 个水平, 每个处理下有 4 个被试。完全随机设计是一种被试间设计, 这意味着每一个被试只接受一种处理或处理的组合³, 故对于本例而言需要 16 个被试。

对接下来要描述的模型的字母进行一下简要的说明:

(1) 假设因素 A 有 p 个水平, 用 α_j 表示其第 j 个水平 ($1 \leq j \leq p$), 也就是说 A 的水平分别是 $\alpha_1, \dots, \alpha_j, \dots, \alpha_p$,

(2) p 个处理下的被试数都相等, 均为 n 个, 用 i 表示每个处理下第 i 个被试 ($1 \leq i \leq n$)

(3) 假设被试在没有施加处理的时候, 无限次实验结果的平均数是 μ 。该值是个理想值, 其实意思是被试测量值在理论上的基线, 不过这么说只是表达了概念。虽然说明了更加明确的定义不过可以充分的感

受到, 这个理论基线值是无法获得的, 毕竟无限次是个极限, 不是一个具体的数, 所以 μ 是个理论值。

(4) 假设实验中的误差是 $\varepsilon_{i(j)}$ ⁴, 其平均数为 0, 方差为 σ_e^2 , 我们进一步假设它服从正态分布

4: 关于为什么它的下标是带括号的现在不便解释。

(4) 观察值 Y_{ij} 指第 j 个处理水平下, 第 i 个被试的观测值

后现的章节因素多了会加入一些新字母和下标, 和上述符号已经没有本质区别, 故以后我不打算再介绍新符号了。

单因素完全随机设计的原假说是: 多个处理水平上的总体平均数相等, 即

$$H_0: \mu_1 = \mu_2 = \dots = \mu_p$$

或处理效应等于为 0, 即

$$\alpha_j = 0$$

单因素完全随机实验设计模型

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{i(j)} \quad (2.1)$$

其中:

Y_{ij}	——	被试 i 在处理水平 j 上的分数
μ	——	总体平均数
α_j	——	水平 j 的处理效应
$\varepsilon_{i(j)}$	——	误差效应

单因素完全随机实验设计与计算举例

研究的问题

我想研究文章的生字密度对学生阅读理解的影响。我的假设是: 阅读理解随着文章中生字密度的增加而下降。因此, 该实验有一个自变量——生字密度, 通过研究前人文献我感兴趣的是四种生字密度 (1)5:1(a_1) (2)10:1(a_2) (3)15:1(a_3) (4)20:1(a_4) 因变量是被试的阅读理解分数。实施实验时, 我将 32 名被试随机分成四组, 每组被试阅读一种生字密度的文章, 并回答阅读理解测验中有关文章内容的问题。这是一个典型的单因素完全随机设计。虽然我不再检验实验中其他因素的影响, 但实际上存在着诸多对因变量影响的无关变量, 比如文章的长度、文章的主题熟悉性、文章类型等, 还有被试的年龄、受教育程度、阅读能力等。这时, 控制无关变量可做的工作之一是在选取四篇文章时, 使它们在除生字密度以外的其他方面尽量匹配。

新的计算模型的引入以及数据的计算

我将引入一种新的计算模式, 这种模式主要是为了手算。为了理解愿意, 我们必须手算一次每种实验设计的简单数据, 但是用统计教材上的算法根本无法实现一个较为精确的估算, 其原因是, 和方的计算公式选用:

$$SS = \sum_{i=1}^n (X_i - \bar{X})^2$$

式子中 \bar{X} 很大可能是小数，且可能是循环小数，所以每次相减都会四舍五入，然后平方后进一步放大误差，这样算到最后误差不断累积。此时要提出的思路是，最后一步再做除法。

Table 2.7: 单因素完全随机实验的 AS 表

	a_1	a_2	a_3	a_4	
	3	4	8	9	
	6	6	9	8	
	4	4	8	8	
	3	2	7	7	
	5	4	5	12	
	7	5	6	13	
	5	3	7	12	
	2	3	6	11	
总	35	31	56	80	202

1. 计算表

2. 各种基本量的计算

$$\sum_{i=1}^n \sum_{j=1}^p Y_{ij} = 3 + 6 + 4 + \dots = 202.000$$

$$\frac{\left(\sum_{i=1}^n \sum_{j=1}^p Y_{ij} \right)^2}{np} = [Y] = \frac{(202)^2}{(8)^2 (4)^2} = 1275.125$$

$$\sum_{i=1}^n \sum_{j=1}^p Y_{ij}^2 = [AS] = (3)^2 + (6)^2 + \dots = 1544.000$$

$$\sum_{i=1}^n \frac{\left(\sum_{j=1}^p Y_{ij} \right)^2}{n} = [A] = \frac{(35)^2}{8} + \frac{(31)^2}{8} + \dots = 1465.250$$

3. 平方和分解与计算

$$SS_{\text{总}} = SS_{\text{组间}} + SS_{\text{组内}}$$

$$SS_{\text{总}} = [AS] - [Y] = 268.875$$

$$SS_{\text{组间}} = SSA = [A] - [Y] = 190.125$$

$$SS_{\text{组内}} = SSE = SS_{\text{总}} - SS_{\text{组间}} = 78.750$$

Table 2.8: 单因素完全随机实验的方差分析表

变异源	SS	df	MS	F	p
组间 (生字密度)	190.125	$p - 1 = 3$	63.375	22.533	< .001
组内 (个体误差)	78.750	$p(n - 1) = 28$	2.813		
总计	268.875	$np - 1 = 31$			

4. 方差分析表及结果的解释 方差分析结果表明，生字密度的效应是显著的 ($F(3, 28) = 22.53 (p < 0.01)$)，生字不同的文章对学生的学生阅读理解有显著的影响。从阅读理解的四种生字密度文章的平均数可以初步看出

5. 平方和与自由度分解图

$$SS_{\text{总}} \begin{cases} SS_{\text{组间}} \\ SS_{\text{组内}} \end{cases} \quad (2.2)$$

$$SS_{\text{总}} \quad df = np - 1 = 31 \quad SS_{\text{组间}} \quad df = p - 1 = 3 \quad SS_{\text{组内}} \quad df = p(n - 1) = 28$$

6. 对平方和分解与计算的一些解释

2.2 单因素随机区组实验设计

2.3 单因素拉丁方实验设计

2.4 单因素被试内实验设计

Greek letters with pronunciation

Character	Name	Character	Name
α	alpha <i>AL-fuh</i>	ν	nu <i>NEW</i>
β	beta <i>BAY-tuh</i>	ξ, Ξ	xi <i>KSIGH</i>
γ, Γ	gamma <i>GAM-muh</i>	\omicron	omicron <i>OM-uh-CRON</i>
δ, Δ	delta <i>DEL-tuh</i>	π, Π	pi <i>PIE</i>
ϵ	epsilon <i>EP-suh-lon</i>	ρ	rho <i>ROW</i>
ζ	zeta <i>ZAY-tuh</i>	σ, Σ	sigma <i>SIG-muh</i>
η	eta <i>AY-tuh</i>	τ	tau <i>TOW (as in cow)</i>
θ, Θ	theta <i>THAY-tuh</i>	υ, Υ	upsilon <i>OOP-suh-LON</i>
ι	iota <i>eye-OH-tuh</i>	ϕ, Φ	phi <i>FEE, or FI (as in hi)</i>
κ	kappa <i>KAP-uh</i>	χ	chi <i>KI (as in hi)</i>
λ, Λ	lambda <i>LAM-duh</i>	ψ, Ψ	psi <i>SIGH, or PSIGH</i>
μ	mu <i>MEW</i>	ω, Ω	omega <i>oh-MAY-guh</i>

Capitals shown are the ones that differ from Roman capitals.

