

南开大学 2004 年研究生入学考试试题

考试科目：心理学方法

专业：应用心理学

一、名词解释（每题 5 分，共 30 分）

1. 中间变量
2. 天花板效应
3. 实验范式
4. 事后回溯设计
5. 中位数
6. 检验的显著性水平

二、问答题（共 84 分）

1. 算术平均数和几何平均数分别适用于什么情形？（8 分）

2. 有一数列 $t_1, t_2 \cdots t_n$ ，令 $Y = \sum_{i=1}^n (t_i - k)^2$ 。K 为何值时 Y 最小？证明你的结论。（8 分）

3. 什么是组间设计？什么是组内设计？分别举一个适合用组间设计而不适合用组内设计的例子，以及一个适合用组内设计而不适合用组间设计的例子。（12 分）

4. 有人说下图所示的拉丁方实验设计的目的是为了研究自变量 A（分别处在 a1、a2、a3 和 a4 水平）对因变量的影响，自变量 B（分别处在 b1、b2、b3 和 b4 水平）和自变量 C（分别处在 c1、c2、c3 和 c4 水平）处于次要地位。这种说法对吗？为什么？（10 分）

	c1	c2	c3	c4
b1	a1	a2	a3	a4
b2	a2	a3	a4	a1
b3	a3	a4	a1	a2
b4	a4	a1	a2	a3

5. 一研究者想探查自变量 A 和 B 对因变量 Y 的影响，A 有 3 个水平，B 有 4 个水平。当这两个自变量无交互作用时应该怎样设计这个实验？说明这样设计的理由。（10 分）

6. 为什么会发生统计回归现象？举一例说明统计回归怎样导致自变量的混淆。（12 分）

7. 一实验涉及的自变量是 X，因变量是 Y；研究假设是： $Y = fX^k$ （f 和 k 是常数，但其具体数值未知）；实验获得的数据是： $(X_1, Y_1), (X_2, Y_2) \cdots (X_n, Y_n)$ 。如何利用这组数据检验这个研究假设？（12 分）

8. 一研究者随机调查了 8 名看过 A 和 B 这两部电影的受访者，问他们更喜欢哪部电影。6 名受访者喜欢电影 A，2 名受访者喜欢电影 B。他就此得出结论：电影 A 更为观众所喜欢。但实际上也有可能这两部电影为观众同等程度地所喜欢而使得这个结论是错误的，发生这个错误的概率是多少？（计算结果可以用分数表示）（12 分）

三、应用题（共 36 分）

1. 在心理学实验中有可能发生下面的自变量混淆：主试以一种不经意的方式向被试流露出自己对实验结果的期望，从而使得被试按照主试所期望的方式做出反应。设计一个实验揭示这种现象的存在。可以采取一些什么措施消除这种影响？（18 分）

2. 某省进行了一次小学五年级的数学统考。已知不同小学教学水平相差较大，但同一个小学的五年级的不同班级教学水平很相近。以学生的考试成绩为原始数据，问：

(1) 如何处理这些原始数据，使得数据处理的结果能够比较不同小学学生的数学学习

潜能?

(2) 如何处理这些原始数据, 使得数据处理的结果能够反应一个学校的教学水平? (此题共 18 分)

参考答案

南开大学 2004 年研究生入学考试试题

考试科目: 心理学方法

专业: 应用心理学

一、名词解释 (每题 5 分, 共 30 分)

1. 中间变量: 指实验中不同于研究的相关变量, 又叫额外变量、无关变量。它们是除自变量之外, 一切能够影响实验结果的变量, 是必须在实验中加以控制的变量。如果应该控制的变量没有控制好, 那么它就会造成因变量的变化, 在这种情况下, 研究者选定的自变量与一些未控制好的因素共同造成了因变量的变化, 这就叫自变量的混淆。必须加以控制, 否则影响因变量的变化。可以采取排除法、恒定法、匹配法、随机化法、抵消平衡法和统计控制法等加以控制。

2. 天花板效应: 指在测验中由于题目的数量少或难度不够, 使被试的得分受到限制, 通过测验不能充分展示他们的能力, 也叫高限效应。要求被试完成的任务过于容易, 所有不同水平 (数量) 的自变量都获得很好的结果, 并且没有什么差别的情况。例如, 同是 100% 的正确, 或同是 100 分, 各个受测者在此方面的反应并不尽相同, 有的很吃力, 有的很轻松, 有的勉强强, 有的绰绰有余, 但这时都是以一种方式记录。这就是天花板效应的表现。与此相对应的是地板效应。

3. 实验范式: 范式是使一门学科成为科学的必要条件或成熟标志, 也就是把范式看作科学与非科学的分界标准, 是科学家集团所共同接受的一组假说、理论、准则和方法的总和, 这些东西在心理上形成科学家的共同信念。从实验内省法到行为主义, 再到认知主义, 都形成了相对独立系统的实验研究范式。

4. 事后回溯设计: 又叫相关设计, 这种设计既不能操纵、控制自变量, 也不能随机分派被试, 只能将被试的现存差异, 分类归到一定的情景中去。这种设计研究的对象是已发生过的事件。在研究过程中, 研究者不需要设计实验处理或操纵自变量, 只需通过观察存在的条件或事实, 将这种已自然发生的处理或自变量与某种结果或因变量联系起来加以分析, 以便从中发现某种可能的简单关系。主要包括两种类型: 相关研究设计和准则组设计。

5. 中位数: 是位于依一定顺序排列的一组数据中央位置的数值, 在这一数值上、下各有一半频数分布着。在心理测量中, 常通过计算百分位数来说明、解释和评价分数在团体中所处的位置。中位数虽然也具备一个良好的集中量所应具备的某些条件, 例如比较严格确定、简明易懂, 计算简便, 受抽样变动影响较小, 但是它不适合进一步的代数运算。它适用于以下几种情况: ①一组数据中有特大或特小两极端数值时; ②一组数据中有个别数据不确切时; ③资料属于等级性质时。

6. 检验的显著性水平: 在统计研究中我们把用在拒绝虚无假设中的概率称为检验的显著性水平, 并用 α 表示。通常我们选择 $\alpha = 0.05$ 作为检验的显著性水平。也就是说每当实验结果发生的概率小于或等于 0.05 的时候, 我们就拒绝虚无假设。比如方程的显著性水平 α 在统计检验中具有重要作用, $\alpha = 0.05$, 意味着回归方程的有效性为 95%, $\alpha = 0.01$, 为 99% 的可靠性。通常 $\alpha = 0.01$, 为高度显著; $\alpha = 0.05$, 为一般显著; $\alpha = 0.10$ 以上, 方程可靠性大为下降。

二、问答题 (共 84 分)

1. 算术平均数和几何平均数分别适用于什么情形? (8 分)

答：(1) 算术平均数是所有观察值得总和除以总频数所得之商，简称为平均数或均数。算术平均数的优点是：①反应灵敏；②严密确定。简明易懂，计算方便；③适合代数运算；④受抽样变动的影响较小。

除此之外，算术平均数还有几个特殊的优点：①只知一组观察值的总和及总频数就可以求出算术平均数。②用加权法可以求出几个平均数的总平均数。③用样本数据推断总体集中量时，算术平均数最接近于总体集中量的真值，它是总体平均数的最好估计值。④在计算方差、标准差、相关系数以及进行统计推断时，都要用到它。

算术平均数的缺点：①易受两极端数值(极大或极小)的影响。②一组数据中某个数值的大小不够确切时就无法计算其算术平均数。

条件要求：①数据必须是同质的，即同一种测量工具所测量的某一特质。②数据取值必须明确。③数据离散不能太大。

从而可以看出，平时各科成绩相加求总分，并按此排序是有问题的，不满足求平均数的要求。

(2) 几何平均数是指一种由 n 个正数之乘积的 n 次根表示的平均数。在计算学校经费的增加率、平均率，学生入学率，毕业生的增加率的计算时常用。

应用：①求学习、记忆的平均进步率。②求学校经费平均增加率，学生平均入学率、平均增加率，平均人口出生率。

2. 有一数列 t_1, t_2, \dots, t_n ，令 $Y = \sum_{i=1}^n (t_i - k)^2$ 。K 为何值时 Y 最小？证明你的结论。(8分)

答：这是用最小二乘法得到比较精确的回归方程的思路。

用 $\varphi(x)$ 拟合已知数据 (x_k, y_k) ($k=1, 2, \dots, n$)，使得误差的平方和

$\sum_{k=1}^n [y_k - \varphi(x_k)]^2$ 为最小，这种求 $\varphi(x)$ 的方法，就是最小二乘法。

直线拟合，设拟合直线为 $y = a_0 + a_1 x (= \varphi(x))$ ， a_0, a_1 满足法方程组

$$\begin{cases} na_0 + (\sum_{k=1}^n x_k) a_1 = \sum_{k=1}^n y_k \\ (\sum_{k=1}^n x_k) a_0 + (\sum_{k=1}^n x_k^2) a_1 = \sum_{k=1}^n x_k y_k \end{cases}$$

K 表示每一点沿 Y 轴方向到直线的距离的平方和最小，所以有： $K = (t_n - Y_n)^2$ 值最小时， Y 最小。

3. 什么是组间设计？什么是组内设计？分别举一个适合用组间设计而不适合用组内设计的例子，以及一个适合用组内设计而不适合用组间设计的例子。(12分)

答：组间设计是指每个被试只接受一种自变量水平的实验处理，不同的被试接受不同的自变量水平的实验处理。在这种设计中由于被试是随机取样并随机安排接受不同的实验处理的，因而又称完全随机化设计；各实验处理组之间没有相关，因而又称独立样本设计。

组内设计是指每个或每组被试接受所有自变量水平的实验处理的真实实验设计。又称“重

复测量设计”。如果用被试组代替单个被试，每组都要接受所有的实验处理，但组中的每个被试随机地接受一种实验处理，这样的组叫区组，这种设计称为“随机区组设计”。同一区组内的被试应尽量“同质”（即在各方面都相似或相同）。随机区组设计的实验处理之间有相关，因此又称“相关样本设计”。被试内设计将被试的个别差异从被试（组）内变异中分离出来，提高了效率。对于容易导致的练习或疲劳效应，通常采取全随机法、对抗平衡法(ADUA)、拉丁法等方法予以消除。被试内设计可分为完全和不完全被试内设计两种，两者的区别在于每个或每组被试接受所有的还是部分的实验处理。结果可用方差分析处理。

举例：(1) 为了避免相同被试前后测试效应的影响，有的设计只适用组间设计而不适用组内设计，比如：设计一个记忆实验，要求被试在用两种不同的记忆策略回忆相同的一段材料。

(2) 有时需要考察相同被试在不同的测验下的结果，这时就适用组内设计而不适用组间设计。比如：班上的学生作为被试，教师是实验者。实验作业是要求被试估计每段时间的长短，刺激时距有四种水平：5 秒，8 秒，14 秒和 23 秒，被试用划线的长短表示时距长短。

4. 有人说下图所示的拉丁方实验设计的目的是为了研究自变量 A（分别处在 a1、a2、a3 和 a4 水平）对因变量的影响，自变量 B（分别处在 b1、b2、b3 和 b4 水平）和自变量 C（分别处在 c1、c2、c3 和 c4 水平）处于次要地位。这种说法对吗？为什么？（10 分）

	c1	c2	c3	c4
b1	a1	a2	a3	a4
b2	a2	a3	a4	a1
b3	a3	a4	a1	a2
b4	a4	a1	a2	a3

答：拉丁方实验设计是多变量实验设计中一种较为常用的设计方案。心理实验中采用循环平衡实验顺序对实验结果的影响，就使实验顺序、被试差异都作为自变量因素处理。只要这些实验中自变量的个数（因素）与实验处理水平数相同，而且这些自变量之间没有交互作用存在时，都可采用拉丁方设计方案。

a1、a2、a3 和 a4 为实验处理的四种水平，b1、b2、b3 和 b4 为被试的四种类型，c1、c2、c3 和 c4 为实验顺序。其中被试类型存在个体差异，被试可以是一人，也可以是多人。根据这种拉丁方实验设计分析法，可以分析出被试间、实验顺序、实验处理间的差异情形，也就是说实验顺序、被试差异对实验结果的影响都可以分析出来。

因此，题目中的说法是不正确的。

5. 一研究者想探查自变量 A 和 B 对因变量 Y 的影响，A 有 3 个水平，B 有 4 个水平。当这两个自变量无交互作用时应该怎样设计这个实验？说明这样设计的理由。（10 分）

答：交互作用是指在实验设计中，一个自变量在另一个自变量不同的水平上表现出了差异。这时给实验结果提供了更多更真实的信息。

如果在实验中交互作用不显著或无交互作用，可能是因为两个自变量的主效应相互独立。这时需要改进实验设计。

如果需要两个自变量之间表现出交互作用，就是说自变量 A 的 3 个水平分别在自变量 B 有 4 个水平上要表现出差异。那么可以采用将被试先按照不同水平的特质配对，然后随机分组进行实验。通过不同被试的重复实验可以获得两个自变量之间的交互作用。这就是说可以控制其中一个自变量的不同水平，考察其在另一自变量水平上的表现。

6. 为什么会发生统计回归现象？举一例说明统计回归怎样导致自变量的混淆。（12 分）

答：统计回归现象指受试者的测量分数在第二次测量时，有向团体平均数回归（趋近）的倾向。换言之，高分组在第二次测量时，其分数由于向平均数回归而有降低的趋势，但是

低分的受试者，其分数却有升高的趋势。在有前后测的实践中。若以极端分数（高分和低分组）的学生为对象，容易出现统计回归现象。极端分数会有向平均数回归的现象，主要理由是机会误差的因素对极端分数的影响，比对普通分数影响大。

如果应该控制的变量没有控制好，那么，它会造成因变量的变化，这时，研究者选定的自变量与一些没有控制好的因素共同造成了因变量的变化，这种情况就称为自变量混淆。

举例：在一项言语学习实验中，研究者想要比较集中学习（一次学习大量材料）和分散学习（分次学习）的效果。他选择三组学生为被试，三组被试的学习材料相同。实验安排如表所示。表中的数字3，表示学习时间为3小时。所有被试均进行相同的考试。

三组被试实验程序

	星期一	星期二	星期三	星期四	星期五
第一组	3	—	—	—	测验
第二组	3	3	—	—	测验
第三组	3	3	3	—	测验

结果表明，第三组的考试分数最高，第二组次之，第一组最差。因此研究者得出结论：分散学习比集中学习效果好。

我们很容易看到本实验存在着自变量混淆，即被试不仅在学习方式上存在差异，而且在学习时间上也存在明显差异。另外，三组被试学习后与测验之间的时间间隔对测验结果也会产生影响。第三组被试学习后只隔一天就进行测验，因此，他们的保持效果较好，遗忘较少，测验成绩就高。总之，本实验设计把学习方式与学习时间和间隔时间等变量混淆在一起，实验结果出现了偏差。

在具体实验操作时，可以通过实验设计来避免自变量混淆。对于出现的问题，我们可以采取消除额外变量、使额外变量保持恒定或随机化等方法，来避免其对因变量的影响。

7. 一实验涉及的自变量是 X，因变量是 Y；研究假设是： $Y = fX^k$ （f 和 k 是常数，但其具体数值未知）；实验获得的数据是： $(X_1, Y_1), (X_2, Y_2) \cdots (X_n, Y_n)$ 。如何利用这组数据检验这个研究假设？（12 分）

答：这个实验建立了研究假设的函数关系模型，并通过实验获得了自变量和因变量之间的相应数值，这些数值为一一对应关系。

通过测量数据去检验统计模型，实际上就是研究假设的统计推断过程。

统计推断是根据带随机性的观测数据（样本）以及问题的条件和假定（模型），而对未知事物作出的，以概率形式表述的推断。它是数理统计学的主要任务，其理论和方法构成数理统计学的主要内容。统计推断的一个基本特点是：其所依据的条件中包含有带随机性的观测数据。以随机现象为研究对象的概率论，是统计推断的理论基础。在数理统计学中，统计推断问题常表述为如下形式：所研究的问题有一个确定的总体，其总体分布未知或部分未知，通过从该总体中抽取的样本（观测数据）作出与未知分布有关的某种结论。例如，某一群人的身高构成一个总体，通常认为身高是服从正态分布的，但不知道这个总体的均值，随机抽部分人，测得身高的值，用这些数据来估计这群人的平均身高，这就是一种统计推断形式，即参数估计。若感兴趣的问题是“平均身高是否超过 1.7（米）”，就需要通过样本检验此命题是否成立，这也是一种推断形式，即假设检验。由于统计推断是由部分（样本）推断整体（总体），因此根据样本对总体所作的推断，不可能是完全精确和可靠的，其结论要以概率的形式表达。统计推断的目的，是利用问题的基本假定及包含在观测数据中的信息，作出尽量精确和可靠的结论。

检验研究假设可以通过以下一些方式：参数估计、假设检验（如：Z 检验、T 检验、 X^2 检验、F 检验）、样本检验等。

8. 一研究者随机调查了 8 名看过 A 和 B 这两部电影的受访者，问他们更喜欢哪部电影。6 名受访者喜欢电影 A，2 名受访者喜欢电影 B。他就此得出结论：电影 A 更为观众所喜欢。但实际上也有可能这两部电影为观众同等程度地所喜欢而使得这个结论是错误的，发生这个错误的概率是多少？（计算结果可以用分数表示）（12 分）

答：概率是指用来表示随机事件发生可能性及出现某种结果可能性大小的数值。概率分布必须符合以下两个要求：所有的概率是介于 0 与 1 之间的一个数；所有结果的概率之和应等于 1。

这个研究中研究者出现结论错误主要是因为以下几个原因：

(1) 调查的样本过少，这样影响研究的可靠性。从题目中可以看出，这个研究者所选择的样本只有 8 个，而且是同时对两部电影的喜欢程度的评价，通过这样的少数样本就得出结论，显然是有较大的错误几率。

(2) 具体到 8 个被试对这两部电影的喜欢程度来说，受访者存在多种倾向，比如两部都不喜欢，喜欢其中一部，两部都喜欢等维度。8 名被试中就存在了许多交互关系，在考虑被试较少的同时又可以看见，当主试让被试二择一时，被试就会在自己的喜欢程度里面犹豫，出现了随机几率。

因此这个实验者在实验设计中就存在了不足，影响了研究结果。

三、应用题（共 36 分）

1. 在心理学实验中有可能发生下面的自变量混淆：主试以一种不经意的方式向被试流露出自己对实验结果的期望，从而使得被试按照主试所期望的方式做出反应。设计一个实验揭示这种现象的存在。可以采取一些什么措施消除这种影响？（18 分）

答：(1) 这种自变量混淆时来自于主试效应的影响。所谓“主试效应”是指“由于实验者对研究结果的期望而产生的实验偏差”。主试效应常常被称作为“皮格马利翁效应”，传说塞浦路斯国王皮格马利翁在雕塑一座少女像时十分地钟情于她，竟然使得雕像获得生命变成了真实的少女而与他结为伴侣。心理学家罗森塔尔曾经用实验的方法证实了教育中期望效应的作用，由此，主试效应又被称作为“罗森塔尔效应”。

(2) 进行主试效应的控制可以在实验中采用双盲控制。举例：假想的“药品实验”：主试想要了解一种治疗高血压症的新药的疗效，而他的被试全都是志愿者，换一句话说，主试和被试都知道他们身处实验情境之中。将被试随机分为二组，一组被试（实验组）服用的是降血压的新药，而另一组被试（对照组）服用的只是用淀粉制成的和降血压药外形、口味完全一样的安慰剂。但是，每一位被试并不知道自己服用的究竟是降血压药还是安慰剂，或者说，被试并不知道自己是属于实验组还是对照组，这是一“盲”。主试当然必须清楚地了解每一位被试所服用的是什么，但是，在实验过程中不是由主试自己来为被试测量血压。进行血压测量的人虽然也是行家里手，并且了解正在进行的是实验研究，只是他并不知道每一位被测量的对象所取用的是降血压药还是安慰剂，这就是另一“盲”。

(3) 举例：Greenhaff (1993) 以 12 名受试者，采双盲实验设计，随机分派为肌酸组 (5 g Cr + 1 g 葡萄糖) 与安慰剂组 (6 g 葡萄糖)，每天 4 次连续服用 5 天，随后进行 5 回合，每回合 30 次的最大自主等速收缩 (180 O/s) 测试，研究口服肌酸，对反复性最大自主收缩肌力的影响。研究结果显示，肌酸组其肌肉肌酸含量较高，且于测试过程中，最大肌力显著提高、肌力递减率较小。由于肌酸组在运动时的血清中堆积较少的氨 (ammonia)，研究者认为可能是由于 PCr 增加，使 ATP 再合成更有效率，而延迟或减缓疲劳的发生。

因此，“双盲控制”的目的，对于被试而言，就是要尽可能地保证他们在反应变量上的变化确实能够反映实验变量的作用，而不是他们对自己实验身份（实验组成员还是对照组成

员)意识的心理作用;对于主试而言,就是要尽可能地排除其主观意愿对客观地评价实验结果的干扰。

2. 某省进行了一次小学五年级的数学统考。已知不同小学教学水平相差较大,但同一个小学五年级的不同班级教学水平很相近。以学生的考试成绩为原始数据,问:

(1) 如何处理这些原始数据,使得数据处理的结果能够比较不同小学学生的数学学习潜能?

答:根据题目可以知道,在一个省里面,不同小学的教学水平差异很大,这时可以将测得的分数转化为标准分数进行相对意义上的比较。

标准分数是将原始分数与平均数的距离以标准差为单位表示出来的量表。因为它的基本单位是标准差,所以叫标准分数。标准分数可以通过线形转换或者通过非线性转换得到,由此可将标准分数分为两类:

①线性转换的标准分数

根据标准分数的定义,可通过下式将原始分数直接转换成标准分数:

$$Z = (x - X) / S$$

因为标准分数是从原始分数(x)中减去一个恒定值(平均数 X)再除以一个恒定值(标准差 S)得到的,所以这是一种线形转换。有时也把线性转换的标准分数简称作标准分数或 Z 分数。

②常态化的标准分数

当原始分数不是常态时,为了使导出的量表分数呈常态分布,可先把原始分数转化为百分等级,然后从正态曲线面积表中便可得到对应的标准分数。由这种方式所得到的分数就叫常态化的标准分数。在将分数常态化时有一个前提:只有所测特质的分数实际上应该是常态分布,只是由于测验本身的缺陷或取样误差而使分布稍有偏斜时,才能计算常态化标准分数。

T分数: T分数以50为平均数,以10为标准差。

$$T = 50 + 10Z$$

标准九:标准九的全称为标准化九级分制,是一种比T分数粗糙的量表。它以5为平均数,以2为标准差。

(2) 如何处理这些原始数据,使得数据处理的结果能够反应一个学校的教学水平? (此题共18分)

答:根据心理统计处理方法。

①根据测验成绩建立全省小学数学成绩的常模,确定大致的数学水平趋势。

②将全部成绩汇总,计算标准分来将全部分数转化为正态分数,建立全省这次小学数学五年级统考的正态分布曲线。然后根据某个学校小学数学统考成绩的样本统计量,包括平均差、方差等,按照公式计算出结果,查正态分布表,考察不同学校小学数学成绩在正态分布图中的位置。这样可以获得这个学校在这个省这次数学统考中的相对水平,这个结果可以大致反映学校的教学水平。