**2004 年北京师范大学心理学院研究生入学考试试题**

**心理学研究方法**

一、填空题（每空 2 分，共 10 分）

1. 算术平均数的所用是描述了一组数据的____。

2. 在正态分布中，标准差为_____时百分等级数为_____。

3. 误差包括随机误差和___，他会影响___。

4. 在选择题中，增加题目数量会影响____。

二、1. 什么是速度测验和难度测验？两者的区别是什么？

2. 有人说：“t 检验适用于样本容量小于 30 的情况。Z 检验适用于大样本检验”，谈谈你对此的看法。

3. 简述两个你所了解的测验名称及其用途。

4. 学业考试成绩为 x，智力测验分数为 y，已知这两者的 rxy=0.5，IQ=100+15z，某学校根据学业考试成绩录取学生，录取率为 15%，若一个智商为 115 的学生问你他被录取的可能性为多少，你如何回答他？

5. 如果两总体中的所有个体都进行了智力测验，这两个总体智商的平均数差异是否还需要统计检验？为什么?

6.选择统计检验程序的方法时要考虑哪些条件,才能正确应用统计检验方法分析问题？

7. 哪些测量和统计的原因会导致两个变量之间的相关程度被低估？

8. 举例阐述信号监测论在测量感受性方面的优势及其应用领域。

9. 视觉试验的额外变量有哪些？如何有效的控制这些变量？

10. 简述反应时测量技术的基本思想及其在心理学实验研究中的应用。

三、1. 传统心理物理实验方法测量感受性时会产生哪些方面的误差。请分析可能影响获得可靠数据的因素即可采取的措施。

2. 举例阐述在一个 2[组间]*3[组内]的混合实验中采用哪些统计方法对两个因素的数据变化趋势、因素的主效应、交互作用进行详细和深入的分析？如果交互作用显著，应如何进行进一步的分析，并对统计结果进行解释。

3. 什么是常模参照测验和标准参照测验？比较其异同，并举例说明。

**参考答案**

**2004 年北京师范大学心理学院研究生入学考试试题**

**心理学研究方法**

一、**填空题（每空 2 分，共 10 分）**

1. 算术平均数的所用是描述了一组数据的_____。

2. 在正态分布中，标准差为_____时百分等级数为 16。

3. 误差包括随机误差和_____，他会影响_____。

4. 在选择题中，增加题目数量会影响_____。

**答案：**1. 集中趋势 2. 1 3. 系统误差 效度 4. 信度

二、**1. 什么是速度测验和难度测验？两者的区别是什么？**

**答：**速度测验和难度测验是按测验的时间划分的。

（1）速度测验是指在测验过程中限定时间，看在特定时间段里完成任务的速度，因而题目并没有超过被试的能力水平，测的是反应速度。难度测验则是指在测验过程中不限时间，即一般每一题目都有时间去做，但有些题目不见得能做出来，测的是解题的最高能力。

（2）两者的区别是：速度测验限定时间，因而题目并没有超过被试的能力水平，测的

是反应速度；难度测验是不限时间，即一般每一题目都有时间去做，但有些题目不见得能做出来，测的是解题的最高能力。难度测验的功用在于测量被试的程度高低。它的时间限制的标准通常是使95%的被试都有做完测验的机会。测量由易到难排列，以测量被试解决难题的最高能力。

同时，速度测验在于测量被试作业的快慢，它的测题难度相等，但严格限制时间，看规定时间内所完成的测量数量。一般题目都很容易，只要给够时间，每个人都能做完。但是它就是在题目数量多，时间短的情况下，来检测你的反应速度。你的分数的高低，完全依赖你的反应速度。在这类测验里，几乎没有人能全部做完所有的题目；而难度测验包含不同难度的题目，由易到难的排列，有最难的题目，几乎所有的被试都解答不了。但给你充裕的时间，让你有机会做所有的题目，并在规定的时间里做完你会做的题目。所以这类测验就是检测你解答难题的最高能力。

**2．有人说：“t检验适用于样本容量小于30的情况。Z检验适用于大样本检验”，谈谈你对此的看法。**

**答：**t检验、Z检验都是均值检验的方法。

t检验是比较两组均数差别最常用的方法。当样本容量小于30时，样本的差异平均数与差数的总体平均数的离差统计量呈t分布，这时应该采用t检验。理论上，即使样本量很小时，也可以进行t检验。只要每组中变量呈正态分布，两组方差不会明显不同。当n＞30时，t分布趋向于正态，这时如果样本容量接近30还可以采用t检验，但也可以用z检验近似处理。

z检验是使用标准正态分布曲线推算假设由机遇发生的概率的检验方法，适用于大样本(n＞30)的统计检验。根据数理统计的理论，当样本的容量增大时，样本平均数的抽样分布属于正态分布，这就为大样本的统计检验奠定了基础。当n＞30时，t分布接近正态分布，这时也完全不符t分布，根据显著性水平假设，这时需要用z检验。Z 检验一般用于大样本（n>=30）实验的差异程度的检验。

在平均数的显著性检验中，分两种情况，其一是关于样本平均数与总体平均数差异的显著性检验，在总体服从正态分布，总体方差已知的情况下，用Z检验；总体方差未知的情况下，用t检验。其二是平均数差异的显著性检验，在两个总体都服从正态分布，总体方差均已知的情况下，用Z检验（相关样本和独立样本所用统计量不同）；在两个总体都服从正态分布，但是总体方差未知时，用t检验（所用检验统计量方法与两个总体是否独立以及方差是否相等有关）

因此，t检验与Z检验没有绝对界限。这个观点不完全正确。

**3．简述两个你所了解的测验名称及其用途。**

**答：**韦氏量表：是由韦克斯勒编制的，适合于儿童和成人。这个测验可以反映智力的各个方面。韦氏成人量表包括言语量表和操作量表两个部分。11 个分测验中，常识、数字广度、词汇、算术、理解、类同6个分测验构成言语量表，填图、图片排列、积木图案、物体拼凑、数字符号5个分测验构成操作量表。言语和操作量表交替进行。每个测验的原始分不一样。各个分测验可以测查智力的不同方面，比如常识可以反映被试知识的广度、一般学习能力，并可以此评价被试的文化背景；词汇可以考察言语理解能力，与抽象概括能力有关，能在一定程度上指出被试的知识范围和文化背景;填图测验有趣味性,能测查智力的G因素，具有临床意义;图片排列可以测量被试的知觉组织能力、分析综合能力，以及观察因果关系、社会计划性、预期力和幽默感等方面的特征，还可以测量智力的G因素，可作为跨文化的测验。

罗夏墨迹测验：是由精神病学家罗夏编制，属于投射测验。这个测验基于知觉与人格之间有某种关系的基本假设，即个人对刺激的知觉反应投射出该人的人格。由于它采用非文

字的墨迹图形刺激，因此适合不同国家和民族使用。这套测验共有 10 张以一定顺序排列的墨迹图，都是对称图形，内容无意义。通过与被试的提问和回答记分，画出心理图象，进行解释和分析。这个实验通过被试对图形的分析和提问，深入挖掘被试的内在心理过程，达到分析人格特征。

**4．学业考试成绩为 x，智力测验分数为 y，已知这两者的 rxy=0.5，IQ=100+15z，某学校根据学业考试成绩录取学生，录取率为 15%，若一个智商为 115 的学生问你他被录取的可能性为多少，你如何回答他？**

**答：**由 rxy=0.5，可以看出学业考试成绩与智力测验分数的相关不明显。也就是智力测验分数的多少不能作为预测学业考试成绩的较好指标。

智商为 115，由 IQ=100+15z，可以得出 z=1。这个标准分数显示了这个学生在同龄儿童中的相对位置，说明这个学生处于同龄学生构成的常模中一个标准差的位置。大概在 0.3413 的位置，按照正态分布表，其以上还有大约 15.87%的人数。因此，如果某学校根据学业考试成绩录取学生，录取率为 15%，那么这个学生很有可能录取不上。但是由于智力测验只代表某种程度上的智力表现，而且学校的学业测验与智力测验相关系数不大，所以只能作为参考，不能用来计算和预测。应该告诉他不要迷信测验，认真备考，任何可能性都有。

**5、如果两总体中的所有个体都进行了智力测验，这两个总体智商的平均数差异是否还需要统计检验？为什么？**

**答：**当两总体中的所有个体都进行了智力测验,但不能确定两个总体的分布的时候，直接做两个总体智商的平均数差异检验时不合适的。

智力测验中一般可以获得描述性统计数据。描述统计的方法获得了一组数据的集中量数，差异量数和相关量数(常称为样本统计量)，它们仅代表了某一总体中的样本所具有的特征，在进行检验前，我们并不了解样本来自的总体是否具有相同的数值特征(总体中的相应数值称为参数，总体均值记为 μ，总体标准差记为 σ，总体相关系数记为 ρ)。然而，心理研究的目的是要了解样本来自的总体的特征。为此，可以运用参数统计检验法依据样本的特征对总体的特征进行推断，以获得总体的有关特征。

检验两个总体的平均数差异不仅要考虑总体分布和总体方差,还需要注意两个总体方差是否一致，两个样本是否相关以及两个样本容量是否相同等条件。两个总体均值差异的显著性检验是通过来自均值相同的总体的样本平均数差异进行推断的。因此，两个总体均值差异的显著性检验也就是检验两个样本平均数是否来自均值相同的总体。由于两个总体之间有时是相关的，有时是独立的，因此平均数差数的显著性检验也有不同的方法。

**6．选择统计检验程序的方法时要考虑哪些条件，才能正确应用统计检验方法分析问题？**

**答：**选择统计检验程序的方法时需考虑以下条件：

(1)看总体分布是否已知。如果已知，看是不是正态分布。如果已知样本分布为常态分布就可以选择参数检验法，如果总体分布未知就用非参数检验法。

(2)在参数检验中，如果总体分布为正态，总体方差已知，两样本独立或相关都可以采用 Z 检验；如果总体方差未知，根据样本方差，采取不同的 t 检验。 如果总体分布非正态，总体方差已知，根据样本独立或相关采取 z’检验；如果总体方差未知，根据独立和相关采取不同的 z’检验。

（3）根据题目考虑用单侧还是双侧检验。

(4)在非参数检验中，按照两个样本相关和不相关、精度与容量等，可以采用符号检验、秩和检验等方法。

**7．哪些测量和统计的原因会导致两个变量之间的相关程度被低估？**

**答：**影响两个变量之间的相关程度被低估的原因有：

（1）测量原因：测量方法的选择、两个变量测验材料的选择和收集、测量工具的精确性、测量中出现的误差、测验中主试和被试效应、测量的信度和效度、测验分数的解释等。

（2）统计原因：假设是否正确、变量的选择、总体分布状况的辨别、两个样本独立和相关性的检验、两个样本方差齐性的检验、两个变量回归分析的差误、参数和非参数方法的选择（统计分析方法是否正确）等。

**8．举例阐述信号监测论在测量感受性方面的优势及其应用领域。**

**答：**信号检测论是信息论的一个分支，研究的对象是信息传输系统中信号的接受部分。信号检测理论将被试的感受性和辨别力分离出来，是对传统的心理物理学方法的重大突破。

信号检测理论是信息论的分支，应用了信息加工原理。因为人的感官、中枢分析综合可以看作一个信息处理系统，因此可以对它进行分析，这个理论还可以加深人们对感受系统的理解。信号检测理论引入心理学，解决了传统心理研究方法不能解决的问题，把被试的反应倾向和辨别力区分开来。同时实验表明，用传统心理物理法测得的痛阈提高了，并不意味着痛觉感受性的下降，而常常时由于改变了极痛标准而造成。统计决策理论是信号检测理论的数学基础。这个理论的最大优点是可以把操作者的感觉敏感性和反应偏向分开，为研究提供了分析工具。其应用主要表现在以下几个方面：

（1）医学心理学中的应用。

异常症状既可以出现在病人也可以出现在正常人身上，医生最初的任务是做出“是”或“不是”的决断。一部分研究者则关注更具体的诊断问题。对痛阈的新的认识否认了传统认为痛阈的提高是由于痛感觉的减轻所致。事实上，被试的感觉辨别力始终没有多大改变，所改变的仅仅是他的痛阈报告的标准。

（2）工程心理学中的应用。

在复杂的人机关系中，警戒操作是工程心理学中的一个重要问题。警戒是指操作者在相当长时间内，对环境中偶然出现的某种信号的觉察并做出反应的持续准备状态。对警戒衰退所做的信号检测论分析表明，应该把击中概率和虚报概率两者结合起来，还应把感觉敏感性和反应偏向分开处理才能说明警戒下降的真正原因。

（3）认知研究中的应用（短时记忆，再记忆的研究）。

除了在感知觉方面的研究外，信号检测论还可应用于再认记忆研究中，在再认记忆中，被试所面临的操作实际上是检测当前的刺激（可能识记过，也可能未识记过，既可能是信号，也可能是噪音），将它同记忆痕迹进行“匹配”，作出“是”或“不是”的反应，这一操作可以看作是典型的信号检测论问题。

**9．视觉试验的额外变量有哪些？如何有效的控制这些变量？**

**答：**（1）视觉实验的额外变量有：

①影响视觉适应的因素：适应前照明。照明越强或眼的光适应时间越长，完全适应所需要的时间就越长；器质性病变。先天夜盲，暗适应减弱；维生素 A 缺乏，造成夜视盲；年龄因素。个体在 20～30 岁时感受性高，以后有所下降；感官的相互作用。其他感官的影响可以使感受性提高或降低；红色护目镜有利于暗视觉；实验光的波长不同，得到不同的暗适应曲线。

②影响视敏度的因素：不同的亮度会影响视敏度；物体与背景之间的对比度不同，视敏度将受到影响；视网膜不同部位的视敏度也不同；视觉的适应影响视敏度；闪光盲会降低视敏度；练习可以大大提高对目标物的视敏度。

③影响闪光临界融合频率：闪光临界融合频率随光相的强度增高而增高；刺激面积；在视网膜中，杆体细胞和锥体细胞的闪光临界融合频率是不同的。

④影响颜色知觉的因素：距离；角度；光的强弱；反射率；后效。

（2）可以从以下几个方面加以控制：

①刺激材料的合理性。

②仪器选择的精确性。

③刺激的时间合适。

④控制主试影响。

⑤被试的特点（如情绪、动机、状态）。

**10．简述反应时测量技术的基本思想及其在心理学实验研究中的应用。**

**答：**反应时测量技术是指用来测量反应时的一门技术。反应时指刺激作用于有机体后到明显的反应开始时所需要的时间，即刺激与反应之间的时间间隔。刺激进入有机体时并不会立即有反应，而有一个发动的过程，这个过程在有机体内潜伏着，直至到达运动反应器，才看到一个明显的反应。这个过程包括刺激作用于感官、引起感官的兴奋，并将兴奋传到大脑，大脑对这些兴奋进行加工，再通过传出通路传到运动器官，运动器官接受神经冲动，产生一定的反应，这个过程用的时间就是反应时间。反应时包括简单反应时和复杂反应时，可以测量。

反应时技术的应用主要表现在：

（1）反应时在制作量表和实际应用方面的用处。通过反应时可以直接评量感觉的强度；通过反应时可以制作间接量表；在心理学研究和实际应用中有很大用途。

（2）反应时技术的实际运用。

①减法法：视觉编码和听觉编码实验，波斯纳（Posner,1970）通过应用减法反应时间实验，证明了在短时记忆的短暂时间内，存在着视觉的编码。这说明短时记忆中，先出现一个短暂的视觉编码，然后出现听觉编码，所以随着两个字母相继呈现时间的加大，视觉编码效应逐渐消失，听觉编码效应增大，其反应时间也加大，从而缩小了与 A、a 字母对反应时间的差别；句子—图画匹配实验，这一实验是由柯拉克（H.H.Clak）和蔡斯（W.G.CHase）设计的，实验时给被试看一个句子和一个图画，例如“星形在十字之上”要求被试判断二者是否一致并作出反应，记下反应的时间。句子有八种，主语有“星形”和“十字”，谓语有“在之上”和“在之下”、“不在之上”、“不在之下”；心理旋转实验，1973 年库伯（L.A.Cooper）和谢帕德（R.N.Shepard）设计该实验来证明心理旋转的实际存在。实验选取不同的字母和数字（如 R、J、G、2、5、7 等）为实验材料，将这些材料取正面或反面以及六种不同的倾斜度，让被试反应后记录反应时间。

②加法法：短时记忆信息提取实验，让被试先看 1-6 个数字（识记项目），然后再看一个数字（测试项目），要求被试判定该数字刚才是否识记过，按键反应，记下反应时间。通过实验，斯腾伯格从反应时的变化上确定短时记忆提取过程有独立作用的四个因素，即测试项目的质量、识记项目的数量、反应类型和每个反应类型的相对频率。他认为信息提取过程包括相应的四个独立加工的阶段：刺激编码阶段、顺序比较阶段、二择一的决策阶段和反应组织阶段。

③开窗实验。开窗实验能够比较直接地测量每个加工阶段的时间，并且能比较明显地看出这些加工阶段，就好象开窗一样，一览无遗。比如：给被试呈现 1-4 个字母并在后面标上一个数字，例如“F+3”等，四个字母相继出现，由被试自行按键，当呈现“F+3”时要求被试念出字母标上 F 后的第 3 个字母来（是 I）。通过实验看出转换的加工过程。

**三、1．传统心理物理实验方法测量感受性时会产生哪些方面的误差？请分析可能影响获得可靠数据的因素及可采取的措施。**

**答：**传统心理物理实验方法包括极限法、平均差误法和恒定刺激法等。在测验感受性时常常出现常误和系列效应问题误差。

常误就是系统误差。大部分是因为实验顺序不同，如时间、空间、动作等造成。常见的有习惯误差与期望误差、练习误差与疲劳误差、空间误差、动作误差、时间误差等。系列

效应是指根据一个刺激在量上与整个系列的关系，对它产生过高或过低估计的倾向。

传统心理物理实验方法测量感受性时会产生常见的误差:

(1) 极限法的刺激由递减和递增的两个系列组成，每次呈现刺激后让被试报告，回答是否有感觉。刺激的增减应尽可能的小，目的是系统的探求被试由一类反应到另一类反应的转折点，即在多强刺激时由有感觉变为无感觉；或由无感觉变为有感觉。注意递减、递增系列是交替进行的，数量一致；每个系列的起始点也不一样，以免被试形成定势。

极限法中要求被试以口头报告的形式表示，因此容易受到练习、疲劳、习惯因素的影响。采取 ABBA 法或 AB 法可以平衡误差。实验前要先训练被试，使其掌握训练标准，在整个实验中保持一致。

(2)平均差误法是实验者规定以某一刺激为标准刺激，然后要求被试调节另一比较刺激，使后者在感觉上与标准刺激相等。客观上一般不可能使比较刺激与标准刺激完全一样，于是每一次比较就会得到一个误差，把多次比较的误差平均起来就得到平均误差。因为平均误差与差别阈限成正比，所以可以用平均误差来表示差别感受性。

平均差误法要求被试调整比较刺激与标准刺激，所以容易因为主观原因造成动作误差、空间误差和时间误差等，可以采用多层次的 ABBA 或 AB 法平衡。

(3)恒定刺激法中刺激通常由 5-7 个组成，在实验过程维持不变，因而这种方法叫做恒定刺激法，又叫次数法、常定刺激差别法、正误示例法。刺激的最大强度要大到它被感觉的概率达到 95%左右，刺激的最小强度要小到它被感觉的概率只在 5%左右。各个刺激之间的距离相等，确定几个制定值，与最大间距与最小变化不同，恒定刺激法的刺激是随机呈现的，每个刺激呈现的次数应相等。

恒定刺激法所用的刺激数目少，且不需要随时调整刺激的强度，因此测不易随时改变强度的刺激较为方便。同时它的刺激随机呈现，所以可以克服期望误差和习惯误差。但是猜测的可能性比较多。因此，需要作好实验设计。

**2．举例阐述在一个 2[组间]*3[组内]的混合实验中采用哪些统计方法对两个因素的数据变化趋势、因素的主效应、交互作用进行详细和深入的分析？如果交互作用显著，应如何进行进一步的分析，并对统计结果进行解释。**

**答：**（1）举例：2[组间]*3[组内]的混合实验，研究大学生对红、黄、绿三种灯光的反应是否与灯光的强度有关的实验。其中，红、黄、绿是颜色的三个水平，灯光的强和弱是两个水平；选择被试的数量；一个自变量（灯光的颜色）采取组内设计，而另一个自变量（灯光的强弱）采取组间设计。

可以采用回归分析的方法可以观察两个因素的数据变化趋势，通过测量数据建立回归模型，分析数据的变化方向。

以一个自变量的不同水平为主效应可以进行主效应的分析。比如，以灯光的强弱这个自变量的两个水平作为主效应，分析对另外一个自变量，即灯光的颜色（包含三个不同水平，红黄绿）的影响。

对于交互作用的分析，可以分别选取两个变量之间不同水平的交互关系进行一一的分析，也可以做方差分析来分析不同自变量和水平之间交互的显著性关系。

（2）交互作用是指一个自变量产生的效果在第二个自变量的每一个水平上不一样。如果交互作用显著，需要具体讨论自变量不同水平之间的效应。可以做相关分析和多因素分析。变量的交互作用在某些试验设计中难以利用和分析，我们采用建模的方式用计算机作图可以把交互作用直观的描述出来。例如，温度和时间对产品的质量和产量存在交互作用。采用低温、长时间，可使反应缓和，有利用提高产品质量；采用高温、短时间，可加快反应速度，由于节约了批产时间，使月产量增加。也就是说，要想有效地控制生产中的各个因素，必须建立数学模型，从中发现规律，运用规律，才能更好地提高企业的技术管理水平，并为工序

自动化控制提供了科学依据。

**3．什么是常模参照测验和标准参照测验？比较其异同，并举例说明。**

**答：**（1）标准是指在编制测验和解释测验分数时所依据的知识和技能领域，而不是指分数的分界标准。标准参照性测验就是在对测验结果进行评价时不是以常模为标准，而是根据特定的操作标准和行为领域，对个体做出是否达标或达到什么程度的判断，而不考虑他人分数的测验。

这种测验是将被试的分数与某种标准进行比较来解释。这种测验常常用来检验学习效果，看对指定的内容范围掌握得如何或达到某一标准。衡量测验优劣得主要指标是内容效度。

（2）常模是测验分数的总体分部形态，一般用测验分数的平均数和标准差表示。用常模可以确定一个被试测验分数的相对高低，即他在所属群体的能力或知识连续体上的相对位置。常模参照性测验就是以常模为评价测验分数优劣标准的测验，常模被视为测验分数的参照，它关心的不是一个人能力或知识的绝对水平，而是他在所属群体的能力或知识连续体上的相对位置。

这种测验是将一个人的分数与其他人比较，看其在某一团体中所处的位置。也就是把受测者的成绩与具有某种特征的人所组成的有关团体做比较，根据一个人在团体内的相对位置来报告他的成绩。所谓常模就是指具有某种共同特征的人组成的一个群体。

（3）两种测验的相同点：两种测验都是对测验分数的解释。两种测验都可以给被试提供相应的信息。

两种测验的不同点：两种测验是根据对测验分数的解释参照不同划分的。标准参照性测验是将测验结果同事先规定的标准进行比较，对被试个体的分数作出解释；而常模参照性测验是将测验分数参照常模加以解释，也就是将每一个人的分数同团体中的其他人进行比较，这是一种相对的比较。

（4）举例：

高中生升大学的如何考试就属于常模参照性测验。在这种考试中，一方面需要测试学生对书本知识的掌握水平，但同时也是属于选拔性考试，必然有一些学生被筛选下去，而只保留优秀的考生。这种考试的结果是获得该考生在一起参加考试学生中的相对位置，通过相比较来衡量其实际能力，考虑是否通过，并以这个信息考虑录取与否。

会计师职业资格考试属于标准参照测验。这种考试并不要求该考生与同时参加一类考试的其他考生进行比较，通过比较获得考生在团体中的相对位置，比如说排名等。而是以考试大纲为依据，确定试题范围为测试考生是否达到某一级别的要求，从而划定一定的分数作为标准（比如 60 分），达到这个标准的考生即认为是合格的，可以颁发相应的资格证书。因此这种考试关键取决于试题的内容效度，即内容是否能衡量考生的真实水平。