

Hypothesis Testing in Psychological Research

Yu Bai

School of Educational Science, Tianjin Normal University, Tianjin
Email: baiypsy_edu@163.com

Received: Mar. 2nd, 2019; accepted: Mar. 17th, 2019; published: Mar. 25th, 2019

Abstract

Based on the hypothesis testing, this paper introduces two types of errors, hypothesis (null hypothesis, alternative hypothesis) and significance level to understand the hypothesis testing fully. This paper also summarizes the step that how to build hypothesis, the reason caused errors, the relationship between two errors and the method to control errors and to choose significance level. In order to understand and apply the hypothesis testing more accurately in the analysis of psychological data, the statistical technique widely used in the field of psychology is further analyzed.

Keywords

Alpha Error, Beta Error, Null Hypothesis, Alternative Hypothesis, Significance Level

心理研究中的假设检验

白 玉

天津师范大学教育科学学院, 天津
Email: baiypsy_edu@163.com

收稿日期: 2019年3月2日; 录用日期: 2019年3月17日; 发布日期: 2019年3月25日

摘 要

本文依据假设检验的原理, 通过对两类错误(α 错误、 β 错误)、两类假设(原假设、备择假设)以及显著性水平的分析, 对假设检验进行整体的、深入的梳理。本文在精确介绍基本概率知识的基础上, 对假设检验的原理以及假设检验的步骤、原假设和备择假设的选取与建立进行总结, 同时对假设检验中两类错误的原因、相互关系、控制和显著性水平 α 的选择进行了阐述。对假设检验这一在心理学领域得以广泛应

用的统计技术进一步加以梳理，以期在心理学数据分析中更为精确地理解和运用。

关键词

α 错误, β 错误, 原假设, 备择假设, 显著性水平

Copyright © 2019 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着社会的不断发展, 科研的力度与广度不断加深。特别在心理研究中, 统计学占据了极其重要的地位。统计与概率的思想方法得到了广泛应用, 其中假设检验作为统计推断中的重要部分在心理研究中的应用也越来越多, 这就需要对它有一个比较深入的了解, 以便有效应用实际, 准确分析数据, 精确获得结论, 从而大幅度地避免决策失误。即使对于初学者来说, 通过对假设检验的整体掌握, 也会大大提高实际应用中的效率及准确率。

假设检验是统计推断中最重要的一类问题, 其目的是在分析数据的基础上决定是否接受原假设; 其基本任务是在总体的分布函数完全未知或者只知其形式, 但不知其参数的情况下, 对总体参数或者总体形态做出一个假设, 然后利用样本信息来推断总体的某些未知特性(盛骤, 谢式千, & 潘承毅, 2008), 假设检验自引入到心理学研究后, 在对其理解以及运用上仍存在问题(温忠麟 & 吴艳, 2010; 仲晓波, Huang, & 万荣根, 2008), 对这一在心理学领域得以广泛应用的统计技术仍需进一步加以梳理。

2. 假设检验的原理

2.1. 统计假设

假设是科研中广泛应用的方法之一, 即对研究对象做假定性说明, 然后根据事实进行观察, 由已知理论进行分析, 进而由其特征提出未知部分的特征。而在统计中, 它是专指使用统计学术语对总体参数的假定性说明。

我们可以通过一个简单的例子来进行引入: “某团体进行了比奈智力测试, 结果加权平均数为 110, 已知比奈测试的常模 $\mu_0 = 100$, $\sigma_0 = 16$, 问该团体智力水平(并非这一次测验结果)是否确实与常模水平有差异(张厚粲 & 徐建平, 2009)” 对于这个问题就是想运用这一样本测验来推断该团体的智力水平与一般水平的差异, 即要判定这一次的结果和规范达到的水平之间的差异显著。针对这一问题, 我们可以用 μ_1 来表示该团体进行多次测验结果的总平均数, 因此该问题就转换为证实 $\mu_1 \neq \mu_0 = 100$, 这也是我们的研究假设 H_1 。若通过假设检验我们接受了 H_1 , 就认为该团体的智力水平与一般水平之间存在差异, 相反, 若拒绝了 H_1 则表明两者之间无差异。然而, 在统计学中 H_1 的真实与否是不能被直接检验的, 由此, 我们就建立了与此相对立的假设 H_0 , 其中, $H_0: \mu_1 = \mu_0 = 100$, 被称为虚无假设、无差假设或原假设。因为 H_1 与 H_0 是相互对立的, 所以 H_1 叫做对立假设或者备择假设(邱江 & 张庆林, 2003)。

在假设检验中, 我们是通过接受或者拒绝原假设从而来判断我们的研究假设。原假设和备择假设之间互相排斥, 则它们中有且只有一个是正确的(仲晓波, 2010)。即统计推论的出发点就是原假设。但是, 对于原假设的判断是需要充分的依据, 若不具备, 则不可轻易拒绝原假设。正如著名统计学家费舍曾经

说过：“可以说，每一实验的存在，仅仅是为了给事实一个反驳虚无假设(原假设)的机会” (张厚粲 & 徐建平, 2009)。

2.2. 假设检验中的小概率原理

在推断假设的正确性中，我们主要依据人们在实践中广泛运用的小概率原理。该原理具体指小概率事件在一次试验中被认为实际上是不会发生的。在假设检验中，首先假定原假设是真实的，其次如果在 H_0 成立的基础上，推算出它发生的概率属于小概率事件，然后就会推翻原假设，而接受备择假设 H_1 。相反，若没有足够的把握拒绝 H_0 ，则接受原假设。这种思想就是概率性质的反证法。

然而，多小才算小概率事件呢？我们都是针对具体问题而定的。在大多数的情况下，0.1被认为是小概率，其次由实际需要还有0.05以及0.01这两个标准值，这都是人为规定的。另外，由于一个概率为 p 的事件在 n 次重复试验中发生至少一次的概率为 $1-(1-p)^n$ ，随着 n 的增长，不论 p 值的大小，它都将趋于1。因此，小概率原理中的一次试验若改成多次试验就不合理了(毕庆雨, 1998)。这也提醒我们，相信小概率事件的合理性是统计推断的前提。

2.3. 假设检验的步骤

一个假设检验过程的完整步骤，由五方面的内容构成，以下是其具体分析步骤：

第一：根据问题要求，提出原假设和备择假设。在这里，我们将从平均数的显著性检验来进行探讨。其假设检验有双侧检验单侧检验的左侧检验以及右侧检验三种类型。

第二：选择适当的样本统计量。我们进行的假设检验就是根据样本信息去推断总体的某些特性，但是，若直接用样本的原始观测值去推断，那么这是非常困难的，因为样本数据的容量有时会过于庞大而无法考虑到每一个样本信息。所以，我们可以根据样本构造出统计量，然后借助于该统计量检验假设。同时还得注意根据不同的问题选择适当的检验统计量。

第三：规定显著性水平 α 错误。这里所说的显著性水平其实代表了在假设检验过程中可能会犯的错误之一。具体来说就是指当原假设正确时，却错误性的拒绝了它。我们可以用错误拒绝来进行概括。对于 α 的大小虽然说是人为规定的，但是也有一定的依据(胡竹菁 & 戴海琦, 2017)。如果对于研究问题而言，需要得到的结果比较精确，即更加谨慎的接受虚无假设，那么 α 就应该选取的小一些；相反，若是没有这样的要求， α 就可以稍微大一些，可见 α 的大小对于假设检验的结果确实有一定的影响。

第四：计算检验统计量的值。计算统计量的具体值得根据样本的确切资料，这也就是将收集到的样本数据进行整理的过程。

第五：做出决策。具体来说指的是把第四步检验统计量的具体值和临界值作比较，而后做出接受或者拒绝虚无假设的决定。而临界值的确定是根据第三步的 α 大小和第二步的统计检验量的分布，查找相应的统计表得到的。

即使在实际中需要进行假设检验的各类问题繁多，但是具体的可以分为两类：两个样本统计量之间的差异以及样本统计量与相应总体参数的差异。

3. 原假设和备择假设的选取与建立

通过抽样的方法，建立两类假设，再利用样本所给定的信息推断原假设的正确性，做出相应的决策。因此，原假设的建立应该与具体的研究问题及其研究目的等相联系。通常来说，建立原假设应该遵循以下原则：往往把有把握的、不能轻易被否定的命题作为原假设，而把无把握的以及不能轻易肯定的命题作为备择假设。

当我们的目的是希望取得对某一陈述强有力的支持时，把这一陈述的对立面作为原假设 H_0 ，就如之前的提及， H_1 的真实性是不能被直接证实的，因此才会通过 H_0 的拒绝与否来推断出错误与否。当我们希望得到某一陈述时，它只能通过它的对立面得到间接的证实。若研究者忽略其他陈述而只是关心某一陈述是否成立，那么此时就应该把该陈述直接设为原假设(孙艳, 2013)。

综上，如果从假设检验的理论上来看，对于原假设和备择假设是同时提出的，意味着两者的地位是相同的。但是在遇到具体研究问题时就会发现，这种地位的同等性被打破了——我们做出拒绝原假设的决策是要有充分的根据，也就是说接受 H_0 ，并不因为我们有充分的依据接受它，而仅是由于没有充分的根据拒绝它，这表示原假设 H_0 始终处于被保护的地位，从而使得它不会被轻易否定(杨少华 & 杨林涛, 2009)。而备择假设 H_1 往往是研究者希望证实的假设，拒绝或者接受它，完全依赖于原假设的正确与否。因此，在显著性检验中，选用哪个假设作为原假设 H_0 不是一个数学问题，要视考察问题的目的和要求、视犯两类错误将会带来的不同后果而确定，它依赖于科学背景、惯例和方便性。通常情况下，在对参数作假设检验时，将把相等的、无差别的以及等号成立的命题作为原假设；将待判定、待证明的、不相等的或是有差别的命题作为备择假设。即对于参数假设检验问题，原假设一定是“等于”、“大于等于”以及“小于等于”某值这三种情况。

4. 假设检验中的两类错误

4.1. 两类错误的产生的原因及其概念

假设检验的最终结果是由计算出来的统计检验量的具体值和临界值的比较来确定的，也就是说，只要统计检验量的具体值所处的位置没有落入总体分布中的拒绝域中，就做出接受原假设的决策。然而有一种情况，如果拒绝域面积很小，而统计检验量的具体值刚好落入其中，此时我们就认为小概率事件发生了，那么就有充分的理由否定虚无假设。

如果以平均数显著性检验为例，在双侧检验且 $\alpha = 0.05$ 的情况下，可以知道，拒绝域占总体分布面积的 5%，换句话说，仍然有 5% 的可能使得我们随机抽取的样本落入到拒绝域中，即使此时的 H_0 为真，我们也不得不做出拒绝 H_0 的决策。此时 5% 的错误决策就称作 I 类错误。又叫做拒真错误，简记做 α 错误，但是从习惯上来讲， α 代表这某一个显著性水平。另一种情况，则与 α 错误相反。当统计检测量的具体值没有落入拒绝域中，我们要做出接受 H_0 的决策，仍然会犯错误。按照之前的例题来说，有 5% 的概率拒绝 H_0 ，但是因为 H_0 处于接受区间中，则有充足的理由接受 H_0 。即此时的 H_0 为假，却错误的接受了它。这就就是所谓的 II 类错误，又称为取伪错误，简记做 β 错误。无论是拒绝还是接受原假设，都有犯错误的可能，所以不能 100% 的得出样本与总体或者样本与样本之间完全无差异的结论。

4.2. 两类错误之间的关系

通过之前的分析，可以得知， α 和 β 其实是两个完全不同的概率，其原因主要是因为两者前提不同。但是，从另一方面来讲，两者之间又并非是完全无关系的。我们主要从下面几点进行讨论：

通过之前的分析，可以得知， α 和 β 其实是两个完全不同的概率，其原因主要是因为两者前提不同。但是，从另一方面来讲，两者之间又并非是完全无关系的。我们主要从下面几点进行讨论：

首先 $\alpha + \beta$ 不一定等于 1，如图 1 所示(张厚粲 & 徐建平, 2009)，左边表示 H_0 为真时 \bar{X}_i 的总体正态分布，而右边则表示 H_1 为真时 \bar{X}_i 的总体正态分布。当在规定的 α 值下，查找出接受或者拒绝的临界值 \bar{X}_α ，因此，在 H_0 为真的情况下，正确接受 H_0 则为 \bar{X}_α 的左边区域，概率大小为 $1 - \alpha$ ；做出错误接受 H_0 的决策概率是 α ，即 \bar{X}_α 右边的区域面积，此时所犯的错误则为 I 类错误。因此，正确接受与错误拒绝 H_0 是在同一前提下的两个概率，它俩之间的概率和必然为 1。比如当 $\alpha + 0.01$ 时，犯错误的可能性为 0.01，而正

确定的可能性为 0.99。

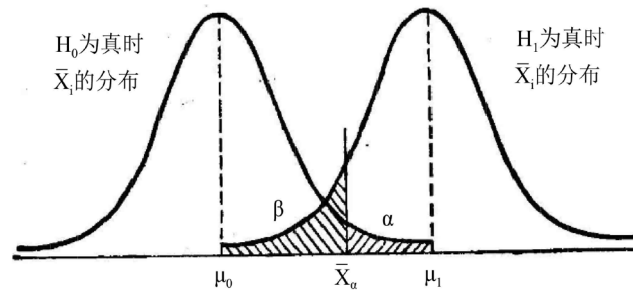


Figure 1. Relation of α and β

图 1. α 和 β 关系示意图

但是当讨论 β 的大小时，所假设的前提已经不是 H_0 为真，然而，由于对 H_0 真假的确定性事先是不知道的，所以此时就可选取它的对立面，即 H_1 为真。在相同 α 大小的情况下，当发生 II 类错误时，就表示错误地接受了 H_0 （错误拒绝 H_1 ）， \bar{X}_α 的左侧就代表了这一情况——当 \bar{X}_i 落入此区域时，做出了拒绝 H_1 的决策，却与前提 H_1 为真不相符合。因此 \bar{X}_α 的左侧区域表示了 β 的概率。从图中可以明显的看出， $\alpha + \beta$ 的总和并不是 1。在一种极端情况下， $\alpha + \beta$ 的总和可以为 1：当我们完全不犯 I 类错误时，即 $\alpha = 0$ ，此时的拒绝区域将成为空集，接受域则为全部的样本空间，对于在 H_1 为真的分布中，则所有的样本都该被拒绝，此时也就意味着 $\beta = 1$ 。

另一方面，在其他情况不变的情况下， α 和 β 不可能同时增大或者减小，针对上图也可以看出此结论。其他情况不变时(包括样本容量 n 等)，若是改变 α 的大小， \bar{X}_α 的大小也会在图形中左右移动，这里就会有两种情况：其一，当 α 减小时， \bar{X}_α 向右移动， β 的值就会变大；其二，当 α 增大时， \bar{X}_α 向左移动， β 的值就会变大；这就表明，在同一问题中(其他情况不变时)，特别是在样本容量 n 确定后，若想同时且同方向控制 α 和 β 是不可能的。

对于假设的确定就是通过找到充分的理由拒绝 H_0 ，从而证实我们之前想得到的假设 H_1 ，而通过以上的描述，可以看出 α 对于最终想得到证实的假设的正确性起到非常关键的作用(焦璨 & 张敏强, 2014)，所以， α 大小的控制必须严格，必要时可以选取 $\alpha = 0.01$ 。

4.3. 两类错误的控制

通过对两者关系的分析得知， α 和 β 在其他情况相同时，它们的大小是具有朝相反方向变化的趋势。但这并不代表我们就没有办法做到同时减小两者的值。一般通过增大样本容量 n ，尽可能的使用单侧检验或根据研究的需要选择适当的 α 水平三方面来进行两类错误的控制。但使用单侧检验时需注意，由于单侧检验本身的性质，会人为性地把所要研究的无方向问题变为单方向问题，因此对于这一种方法，应该谨慎使用。在其他条件都不变的情况下， α 和 β 之间是此增彼减的关系。例如在研究两个样本之间的差异显著性时，我们要有足够的把握拒绝原假设从而使得“两个样本之间存在差异”这个结论有更高的准确性，这样就要求尽可能的减少 α 错误，使犯弃真错误的概率降低。当然，若当一个研究力求证明所比较的两个样本在某个指标上没有差异时，意味着得提高接受 H_0 准确性，反言之，就得降低 β 错误，因此可将 α 水平调大以减小 β 水平，以提高统计结论的可靠性(吴艳 & 温忠麟, 2011)。综上所述，面对实际问题时，应该根据研究需要，选择适当的 α 大小来避免相应的错误。

5. 显著性水平 α 的选择

对于一个显著性水平为 0.05 的假设检验，当想对总体均值 μ 的取值有一定了解时，总是先做出一个

假设,即零假设 $\mu = \mu_0$ 。在该假设成立的条件下,样本均值 \bar{X} 就会确定出一个唯一特定的抽样分布,当 $\alpha = 0.05$ 时,就会假设 \bar{X} 落在对应区间的概率为0.95。根据小概率原理,在零假设成立时,即使是一次抽样的样本 \bar{X} ,它落在该区间以外的概率也很小,可能性几乎等于0。由上可知,在一次抽样后, \bar{X} 会出现两种结果,对于这两种结果存在不同的关于总体均值 μ 的结论:如果 \bar{X} 落入否定域,我们就会在否定零假设的同时断言 $\mu \neq \mu_0$;如果情况相反,则不否定零假设,并且不做出关于总体均值的任何结论。

在心理学研究中显著性水平的意义都是在一个前提下进行的:零假设为真,在这个条件下,我们可以依据一次随机抽样后的样本均值 \bar{X} 与否定域的包含关系来决定是否否定零假设。在这个随机试验中,只有样本自身或者依据于样本值所得出的结果是随机的,即样本、样本均值 \bar{X} 、 \bar{X} 是否落入否定域、是否否定零假设、是否产生错误结论则是随机的。而最初所做出的零假设以及根据抽样分布而得出零假设为真以及做出断言的方式是确定的,因此显著性水平是指零假设为真的条件下假设检验这种断言方法犯错误的概率;这也就是说,它是针对假设检验这种方法,而与由此方法已经获得的结论的错误概率无关;就如之前所分析的,在这个条件下,通过假设检验的一系列步骤所得到的结论无非是两种,要么是否定零假设,但是在这种情况下,所得到的结论又必定是错误的;要么是无法得到任何结论。在另一种情况下,即 $\mu \neq \mu_0$,样本均值 \bar{X} 仍有一个相应的抽样分布(H_1 为真时的分布)。若将两个分布放置在一起,就会发现 \bar{X} 落在零假设所划定的否定域的概率大于0.05,落在否定域之外的概率小于0.95。通过假设检验,一方面我们会选择否定这个错误的零假设,断言 $\mu \neq \mu_0$,这时候也便得到了正确的结论;但另一方面也可能形不成关于总体参数的任何结论;不过在这种情况下,我们不会得出错误的结论。

实际上的总体均值 μ 与 μ_0 的值是否相等是随机的,即我们所提出的零假设与实际情况可能是一致的也可能是相违背的。尽管这两种情况的出现是不确定的,但在处理时,我们 μ 是否等于 μ_0 看成是既定的。这也就是说,在假设检验中总体均值 μ 以及我们选定的 μ_0 以及 μ 是否等于 μ_0 是既定的,没有 $\mu = \mu_0$ 及 $\mu \neq \mu_0$ 各自概率是多少的说法。

6. 结论

综上所述,显著性水平具有以下几方面的含义: α 代表一种犯错误的概率,具体来说,它是在原假设为真的情况下,即样本来自于总体,根据假设检验所判断出来的总体参数取值否定了原假设、推断出了总体的某些特征以及犯错误的概率。另外一方面,它也是在原假设为假时,即实际情况与原假设不相符合,假设检验正确拒绝零假设、并产生正确结论的最小概率。在零假设真伪无法确定,须结合考虑各种情况时,它就是假设检验否定零假设和产生结论的最小概率,以及犯错误的最大概率(Thompson, 2004, 2011)。它不是指假设检验犯错误的概率,不是指假设检验否定零假设、产生结论的概率,也不是指假设检验所产生的结论是错误结论,即否定零假设时错误否定的概率。在假设检验中不存在这些概率。它是针对假设检验这种方法而言的,是对这种推论方法本身性质的描述,不是针对假设检验所产生的结论。

假设检验在心理学研究领域等方面占有重要的作用,本文通过对假设检验的理论分析,可以进一步推动其在具体科研数据分析中的应用,同时对假设检验的精确掌握也更有利于实验得出更为准确的结论。

参考文献

- 毕庆雨(1988). 关于假设检验中的双侧检验与单侧检验. *河北林果研究*, 3(1), 83-92.
- 胡竹菁, 戴海琦(2017). 心理学实验研究的效果大小. *心理学探新*, 37(1), 70-77.
- 焦璨, 张敏强(2014). 迷失的边界: 心理学虚无假设检验方法探究. *中国社会科学*, No. 2, 148-163.
- 邱江, 张庆林(2003). 假设检验策略研究进展述评. *西南师范大学学报(人文社会科学版)*, 29(4), 38-41.
- 盛骤, 谢式千, 潘承毅(2008). *概率论与数理统计*(第4版). 北京: 高等教育出版社.

- 孙艳(2013). 统计假设检验中原假设 H_0 和备择假设 H_1 的探讨. *教育教学论坛*, No. 52, 93-94.
- 温忠麟, 吴艳(2010). 屡遭误用和错批的心理统计. *华南师范大学学报: 社会科学版*, No. 1, 47-54.
- 吴艳, 温忠麟(2011). 与零假设检验有关的统计分析流程. *心理科学*, 34(1), 230-234.
- 杨少华, 杨林涛(2009). 参数假设检验中原假设与备择假设的交换问题. *统计与决策*, No. 5, 148-149.
- 张厚粲, 徐建平(2009). *现代心理与教育统计学*第3版, 页 226, 434-435). 北京: 北京师范大学出版社.
- 仲晓波(2010). 零假设检验和元分析之间的逻辑连贯性. *心理科学*, 33(6), 1477-1480.
- 仲晓波, Huang, X.R., & 万荣根(2008). 心理学中对假设检验一些批评的分析. *心理科学*, 31(4), 1010-1013.
- Thompson, B. (2004). The "Significance" Crisis in Psychology and Education. *The Journal of Socio-Economics*, 33, 607-613. <https://doi.org/10.1016/j.socec.2004.09.034>
- Thompson, B. (2011). Role of Effect Sizes in Contemporary Research in Counseling. *Counseling and Values*, 50, 176-186. <https://doi.org/10.1002/j.2161-007X.2006.tb00054.x>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2160-7273, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: ap@hanspub.org