

水产科学研究中被忽视的II型错误

王迎宾¹, 俞存根^{1*}, 陈勇²

(1. 浙江海洋学院水产学院, 浙江 舟山 316022;

2. 美国缅因大学海洋科学学院, 缅因 04469)

摘要: 显著性检验利用样本信息来判断对总体所做假设的合理性, 检验结果可能犯2类错误, 即I型错误和II型错误(分别为弃真错误和纳伪错误)。随着水产科学研究的不断规范与深入, 显著性检验方法的使用越来越频繁。在国内6个水产期刊近5年(2011年1月-2015年4月)发表的3 829篇学术文章中, 使用了显著性检验方法的有2 235篇, 占到58.4%。然而, 这些显著性检验均只关注了I型错误, 完全忽略了对II型错误的分析。事实上, II型错误并非不重要, 犯该错误的概率也往往会大于I型错误, 而且有时其带来的危害比I型错误更加严重。在水产科学研究中同样如此, 对II型错误的忽视, 很可能导致管理失误, 从而引起渔业资源衰退。本研究对国内外水产科学研究中II型错误分析的应用进行了总结, 阐明了国内水产科学研究中有关II型错误分析缺失的现状; 通过实例, 简要介绍了犯II型错误概率 β (或检验效能 $\text{power} = 1 - \beta$)的计算方法; 分析阐述了影响检验效能的几个主要因素与 β 之间的关系; 针对II型错误分析, 探讨今后水产科学研究与管理中需要注意的问题, 并对其未来发展进行了展望。

关键词: 水产; 显著性检验; II型错误; 检验效能和I型错误

中图分类号: S 91

文献标识码: A

随着水产科学研究的日益深入, 数理统计方法越来越多地应用到了水产科学研究的诸多领域中^[1]。其中, 变量之间的差异显著性检验是重要内容之一, 使用方法主要包括 t 检验、 χ^2 检验和 F 检验等等。这些方法都是教科书中经典的检验方法, 计算的过程都符合传统程序, 因此无论哪种方法都可能会犯统计学中的2类错误: I型错误和II型错误。I型错误也被称为弃真错误, 表示真实情况下两者差异不显著(原假设 H_0 成立), 但检验结果为差异显著(拒绝了 H_0 , 接受备择假设 H_A)。犯I型错误的概率不会超过所选择的显著性水平 α 。II型错误也被称为纳伪错误, 表示真实情况下两者差异显著(H_0 不成立), 但检验结果却是差异不显著(接受了 H_0), 此时犯II型错误的概率为 β ^[2](表1)。

虽然在水产科学研究中, 统计检验方法的使用越来越频繁, 但是一个普遍性的问题是这些检验都只关注I型错误, 即 H_0 被拒绝时, 研究者将做出差异显著的结论, 并指出犯I型错误的概率小于等于 α 。而当检验结果显示两者差异不显著时, 研究者便很自然地接受了 H_0 , 认为两者差异不显著, 但并不再进一步判断此时犯II型错误的大小。因此, 导致在水产科学研究的文献中, 关于此类错误的报道并不多。

事实上, II型错误在水产科学研究中非常重要, 其犯错概率也往往会大于I型错误, 而且其带来的危害是十分严重的, 有时可能比I型错误的危害更大。例如想要分析某种渔业资源品种在适当增加捕捞强度以后, 资源量是否有显著下降? 假设检验的 H_0 为增加捕捞强度后资源量没有显著变化, H_A 为增加捕捞强度后资源量有

收稿日期: 2015-05-28 修回日期: 2015-10-10

资助项目: 国家自然科学基金(31270527); 浙江省公益性技术应用研究计划(2015C33094); 浙江省自然科学基金(LY13D010005); 浙江省本科院校中青年学科带头人学术攀登项目(Pd2013222)

通信作者: 俞存根. E-mail: cgyu@zjou.edu.cn

表1 两类错误之间的关系以及相应概率

Tab. 1 Relationship between two type errors and the corresponding probabilities

检验结果 results of test	未知的真实情况 unknown true situation	
	H_0 成立 H_0 is true	H_0 不成立 H_0 is false
接受 H_0 accept H_0	推断正确(1- α) correct decision (1- α)	II型错误(β) Type II error (β)
否定 H_0 reject H_0	I型错误(α) Type I error (α)	推断正确(1- β) correct decision (1- β)

显著变化。如果检验结果导致I型错误，正确的 H_0 被否定，即捕捞强度增加前后资源量本没有显著差异，但检验者得到却认为资源量有明显下降。根据检验的结果，管理者可能会恢复原始捕捞强度，以保证资源的可持续性，即使事实上增加捕捞强度并没有对资源产生显著影响。相反，如果检验结果导致II型错误，错误的 H_0 被接受，即捕捞强度增加前后资源量有显著差异，但检验者得到却认为资源量有没明显变化。此时，虽然增加捕捞强度事实上对资源会起到破坏效果，但是管理者很可能会根据检验结果，维持或进一步增加捕捞强度。不难发现，在某些情况下，II型错误的成本和危害是高于I型错误的。上述例子中，I型错误会使渔民收入受到些影响，但资源不会面临衰退的风险；而II型错误虽然使渔民的收入短期内增多，但结果却是资源的衰退，以及日后高成本的资源恢复工作。

大多学者忽略对II型错误的分析，可能的原因有以下几点：① β 不像 α 那样直观，即在统计检验过程， α 作为显著性水平直接应用在检验过程中，而这个过程中却并没有出现 β 。② 犯II型错误的概率 β 的大小虽然与 α 相关，但并不能直接读取，计算也相对繁琐。当 H_0 被拒绝时，表明随机误差小于统计上公认的概率水平 α ，此时我们将有1- α 的把握认为 H_0 被拒绝是正确的。相反，当 H_0 未能被拒绝时，表明相互比较的2个总体特征数间的差异在同一总体中出现的可能性大于统计上公认的概率水平 α 。这时，我们将面临犯II型错误的可能。③ 通常人们更看重I型错误，因为一般人们认为 H_0 的提出总是有依据的，在做检验时不能轻易放弃原假设。因此，在统计检验过程中，对I型错误的关注程度往往高于对II型错误的关注。但事实并非如此，很多时候所提出的原假设并没有什么先验的依据，特别是对于水产学这样非确定性很强的学科而言，更是如此， H_0 的选择往往是简单原则。

1 国内外研究概况

在统计检验中，II型错误方面的研究早在20世纪30年代就开始了^[3]。但水产研究领域对该问题的关注则始于20世纪80年代^[4]。最初，这一问题在渔业调查和实验设计中得到关注，学者们通过检验效能(statistical power = 1- α)等的计算，来检验调查和实验设计的合理性，从而提高研究结果的可靠性。

1983年Peterman等^[4]针对美国俄勒冈银鲑(*Oncorhynchus kisutch*)资源中幼鱼和成鱼的关系开展了研究，提出释放 48×10^6 尾幼鱼不足以获得关于幼鱼和成鱼相关关系的变动信息，而若将释放尾数增加至 88×10^6 尾的话则可以更好地对幼鱼和成鱼相关关系做出判断。在研究中，作者计算了检验效能，为使犯I型和II型错误的概率在可接收范围内，提出了以上幼鱼放流策略。随后，Nickelson^[5]对美国俄勒冈银鲑存活率是否与资源密度无关进行了显著性检验，结果为不能拒绝原假设，即存活率与密度无关(成鱼与幼鱼之间符合线性关系)。虽然原假设被接受，但是Nickelson并没有计算检验效能，即并没有衡量犯II型错误概率的大小。针对此问题，Peterman^[6]做了深入分析，发现当初Nickelson所得结果犯II型错误的概率达到了81%。以上是渔业领域对II型错误方面较早的研究报导，研究结果引起了渔业科学家的广泛关注，使渔业研究中检验效能的分析得到重视。1987年，Holt等^[7]对太平洋东部热带海域白点原海豚(*Stenella attenuata*)群体年间变动特征进行检验。他们通过计算检验效能，对调查渔船数量和调查天数进行了分析，指出2艘调查船每年调查120 d，连续调查5年，可以在 α 和 β 都等于10%的水平下探测出白点原海豚年下降率为10% (5年实际总下降率为41%)。即使增加1艘船来加大调查海域面积，对资源量变动的探测结果也并不会会有显著提升

(44%的总下降率中,调查实际探测到的年下降率为11%)。研究指出,当时的调查设计难以有效探测白点原海豚群体资源变动情况。作者在分析中加入了检验效能分析,使得分析结果更加可靠。

1990年, Peterman^[8]对当时渔业领域研究中使用假设检验方法的情况进行了统计,其中接受了原假设却没有做II型错误检验的比例高达98%。研究强调了检验效能的计算在渔业研究中的重要性。通过列举渔业生物学和资源量变动等实例,证实了考虑II型错误出现的可能性能够使渔业研究更科学、渔业管理更有效^[8]。Wagner等^[9]对渔业管理中常会遇到的重要指标(如渔业生物学指标和栖息地等)随时间发生变化的问题进行了综述。该研究指出,检验效能分析是设计渔业监测工作过程中重要的工具。但是,由于渔业管理者所面对的问题通常具有很高的变动性和不确定性,因此,渔业监测目标与适当的统计模型(包含检验效能等)相关联,进而使在可用数据基础上的管理决策是最优的。研究提出, Bayes推断是传统假设检验之外较好的选择,虽然它不一定能够提高研究者对短期内指标变化趋势的判断能力,但它能使我们对趋势的变化不单局限于传统的“差异显著”或“差异不显著”这样的描述。

通过检索汇总,我国在II型错误方面研究的最早报道是郭春彦^[10]在1989年关于教育与心理统

计书中II型错误的分析。研究以高考语文平均成绩为例,重点分析了 α 与 β 的“反比例”关系,以及样本容量 n 与 α 和 β 的关系等问题。随后,张学中^[11]1991年发表论文,从医学角度讨论了 t 检验后II型错误 β 和最小有意义差别 δ_s 的估计问题。自2位学者研究后,从上世纪90年代至今,国内却再未见到有关II型错误的报道。而具体到我国水产领域,虽然各方面的科学研究已经取得了长足的发展,但在统计检验II型错误方面的研究却从未有涉及。我们对国内6个水产类期刊近5年(2011年1月-2015年4月)发表的学术论文进行了查阅,共计3829篇,其中进行了显著性检验的文章有2235篇,但这其中分析II型错误的文章为0篇(表2)。这不仅仅是我国在水产领域研究的缺陷,也是我国在水产领域研究的巨大隐患。

2 检验效能的计算

在此,以一个渔业资源开发与管理的例子来具体介绍检验效能的计算方法。渔业资源管理中,某一资源种类渔获物的平均体长通常作为指示该种类是否遭到过度捕捞的重要指标之一。假设某种渔业资源渔获物平均体长在220 mm以上表示该资源处于合理开发状态,平均体长小于220 mm则表明该资源可能已经遭受过度捕捞。如果几年前该资源群体渔获物平均体长为220 mm,如今想要检验经过几年捕捞后,该资源群体是否处于过度捕捞状态。使用 t 检验,

表2 国内6个水产类期刊近5年发表学术论文使用显著性检验方法及其II型错误分析情况

Tab. 2 Application of significance test method and analysis of Type II error in research articles published in 6 domestic fisheries journals within recent 5 years

期刊名称 journal name	近5年刊出学术论文数 numbers of papers published in recent 5 years	使用显著性检验方法的论文数 number of papers using test of significance approach	进行II型错误分析论文数 number of papers reporting Type II error analysis
水产学报 Journal of Fisheries of China	989	706	0
中国水产科学 Journal of Fishery Sciences of China	638	321	0
渔业科学进展 Progress In Fishery Sciences	480	254	0
水产科学 Fisheries Science	715	433	0
上海海洋大学学报 Journal of Shanghai Ocean University	524	239	0
大连海洋大学学报 Journal of Dalian Ocean University	483	282	0
合计 totals	3829	2235	0

在 $\alpha = 0.05$ 水平下, 检验该资源群体渔获物平均体长是否显著小于220 mm。采集该资源群体渔获样本50尾, 测量得到样本平均体长为218 mm(标准差 $\sigma = 20$)。选择原假设 $H_0: \mu = 220$ mm, 则备择假设 $H_A: \mu < 220$ mm。根据 t 检验结果可知, 差异不显著, 接受原假设 H_0 , 结论为该样本所在总体的平均体长 $\mu = 220$ mm, 即该资源群体并未遭受过度捕捞压力。按照传统的对假设检验结果的判断, 由于接受了原假设, 所以并没有犯I型错误, 因此, 该检验得到的结论应该是可靠的。然而, 不能忽视的一个问题是, 接受原假设意味着会有II型错误发生, 而如果忽视II型错误, 可能给该种类资源群体的保护带来灾难性的结果。即该资源群体渔获物平均体长确实已经小于220 mm, 而通过假设检验却得到了相反的结果, 这样该资源将继续遭受过度捕捞, 从而导致资源的进一步衰退。

针对这个例子, 我们进一步计算可能犯II型错误的概率。 t 检验假设如下:

$$H_0: \mu = 220 \text{ mm}, \quad H_A: \mu < 220 \text{ mm}$$

首先, 计算 $\alpha = 0.05$ 水平下 H_0 拒绝域的临界点, 用 \bar{x}_α 表示。通过查阅标准正态分布表可知:

$$0.05 = P(z < -1.645) \quad (1)$$

$$\text{于是有 } z = \frac{\bar{x}_\alpha - \mu}{\frac{\sigma}{\sqrt{n}}}, \quad (2)$$

$$-1.645 = \frac{\bar{x}_\alpha - 220}{\frac{20}{\sqrt{50}}}$$

$$\text{得到 } \bar{x}_\alpha = 220 - 1.645 \frac{20}{\sqrt{50}} = 215.35 \quad (3)$$

拒绝域为 $x < 215.35$, 即在 $\alpha = 0.05$ 水平下, 只要样本平均数小于215.35 mm, 原假设均将被拒绝(图1)。

相反, 如果原假设 H_0 是错误的, 而备择假设 H_A 正确时, 样本分布为 H_A (图1)。此时, 任何 $\bar{x} \geq 215.35$ 将使得 H_0 被接受。由于 H_0 本身是错误的, 这样 $\bar{x} \geq 215.35$ 必然将导致II型错误的出现。对于以上例子, 测量50尾渔获样本得到平均体长218 mm, 犯II型错误的概率(图1中阴影部分)可计算如下:

$$P(\bar{x} > 215.35) = P(z > \frac{215.35 - 218}{\frac{20}{\sqrt{50}}}) = \quad (4)$$

$$P(z > -0.938) = 0.8236$$

可见, 虽然 t 检验的结果接受了渔获物平均体长

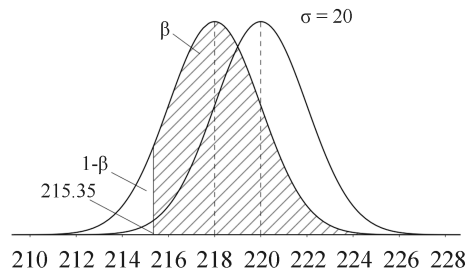


图1 两个抽样分布检验效能 $(1-\beta)$ 示意图

平均值为220和218, 标准差等于20, 显著性水平 α 为0.05, 样本含量等于50。215.35为平均值220的抽样分布的拒绝域临界点。斜线部分为犯II型错误的概率 β 的大小, $1-\beta$ 为检验效能

Fig. 1 Statistical power $(1-\beta)$ for two sampling distributions

Means are 220 and 218, standard deviation is 20, significant level is 0.05 and sample size is 50. 215.35 is the critical point of the rejection region for the sampling distribution with mean of 220. Oblique line area represents the probability of type II error β , and $1-\beta$ equals to power.

大于220 mm的假设, 避免了I型错误, 但II犯型错误的概率却高达82.36%, 检验效能仅为17.64%。

3 影响检验效能的因素

检验效能或 β 的大小并非一成不变, 而是会随着一些因子的改变发生变化, 这些因子主要包括显著性水平 α 、总体标准差 σ 、样本平均数观测值以及样本容量等。

3.1 显著性水平 α 对检验效能的影响

α 和 β 两者是相互牵制的, 当 α 值减小时, β 会增加。例如在上面例子中, 如果其他条件均不变, 仅 α 值从0.05减小到0.01, $0.01 = P(z < -2.327)$ 。此时拒绝域变为 $\bar{x} < 213.42$ (图2)。对平均体长为218 mm(标准差 $\sigma = 20$)的50尾样本进行计算后, 得到 $\beta = 94.42\%$ (图2), 检验效能等于5.58%, 比 $\alpha = 0.05$ 时降低了12.06%。

可见, 一味地减小 α 来达到降低I型错误的目的并不一定是明智的做法, 犯I型错误概率降低伴随的是犯II型错误可能性的增加。 α 不同时($\alpha \in [0.01, 0.5]$), 计算得到的检验效能值的变化趋势(图3)。

3.2 标准差 σ 对检验效能的影响

σ 是表示总体中个体离散程度的参数, 当随机样本的 σ 不同时, 检验效能的大小也会随之发生变化。如果上面渔获物平均体长的例子中,

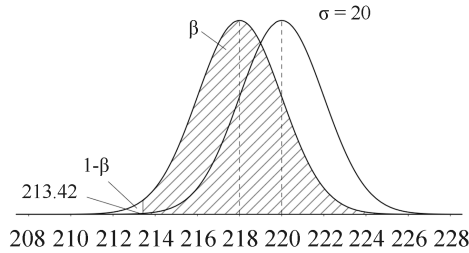


图2 与图1相比显著性水平 α 对检验效能($1-\beta$)的影响情况

平均值为220和218,标准差等于20,显著性水平 α 为0.01,样本含量等于50。213.42为平均值220的抽样分布的拒绝域临界点。斜线部分为犯II型错误的概率 β 的大小, $1-\beta$ 为检验效能

Fig. 2 The effect of significant level α on the statistical power ($1-\beta$) compared with Fig. 1

Means are 220 and 218, standard deviation is 20, significant level is 0.01 and sample size is 50. 213.42 is the critical point of the rejection region for the sampling distribution with mean of 220. Oblique line area represents the probability of Type II error β , and $1-\beta$ equals to power

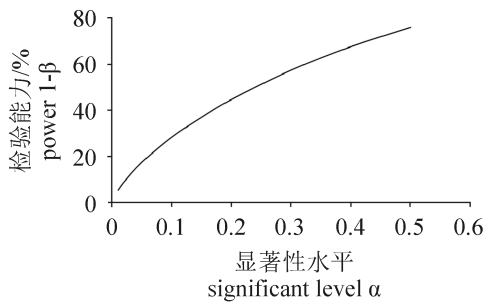


图3 不同显著性水平(α)下检验效能($1-\beta$)值

平均值为220和218,标准差等于20,样本容量50

Fig. 3 Statistical power ($1-\beta$) values for different significant levels α

Means of 220 and 218, standard deviation of 20 and sample content of 50

σ 从20减小到15,其余条件均不改变,则拒绝域变为 $x < 216.51$ (图4)。对平均体长为218 mm的50尾样本进行计算后,得到 $\beta = 75.71\%$ (图4),检验效能等于24.29%,比 $\sigma = 20$ 时增加了6.65%。 σ 减小意味着总体中个体之间差异减小,密度曲线变窄,因此斜线部分面积(即 β 大小)也随之减小(图4)。

3.3 样本平均数观测值 \bar{x} 对检验效能的影响

犯II型错误的概率同样会受到样本平均数观测值的影响。如果同样是50个样本,观测到平均体长变为215.5 mm,标准差不发生变化($\sigma = 20$),则根据公式(4)计算得到犯II型错误的概率将降低至54.28% (图5),检验效能为45.72%,比 $\bar{x} = 218$ mm时增加了28.08%。

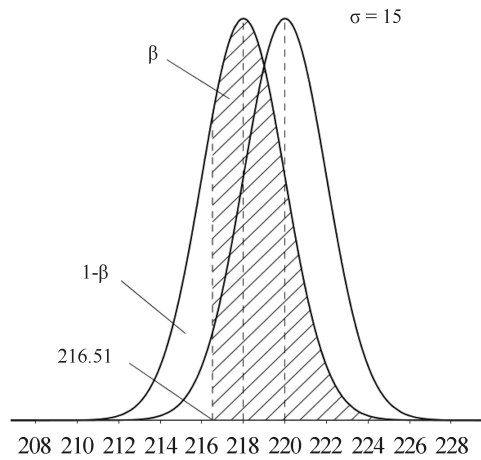


图4 与图1相比标准差 σ 对检验效能($1-\beta$)的影响情况

平均值为220和218,标准差等于15,显著性水平 α 为0.05,样本含量等于50。216.51为平均值220的抽样分布的拒绝域临界点。斜线部分为犯II型错误的概率 β 的大小, $1-\beta$ 为检验效能

Fig. 4 The effect of standard deviation σ on the statistical power ($1-\beta$) compared with Fig. 1

Means are 220 and 218, standard deviation is 15, significant level is 0.05 and sample size is 50. 216.51 is the critical point of the rejection region for the sampling distribution with mean of 220. Oblique line area represents the probability of Type II error β , and $1-\beta$ equals to power

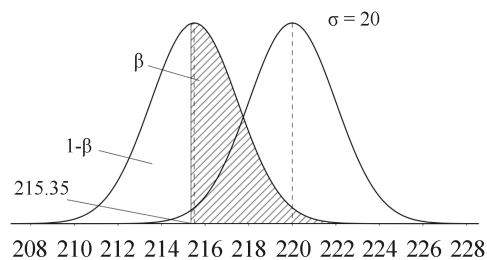


图5 与图1相比样本平均数 \bar{x} 对检验效能($1-\beta$)的影响情况

平均值为220和215.5,标准差等于20,显著性水平 α 为0.05,样本含量等于50。215.35为平均值220的抽样分布的拒绝域临界点。斜线部分为犯II型错误的概率 β 的大小, $1-\beta$ 为检验效能

Fig. 5 The effect of sample mean \bar{x} on the statistical power ($1-\beta$) compared with Fig. 1

Means are 220 and 215.5, standard deviation is 20, significant level is 0.05 and sample size is 50. 215.35 is the critical point of the rejection region for the sampling distribution with mean of 220. Oblique line area represents the probability of Type II error β , and $1-\beta$ equals to power

3.4 样本容量对检验效能的影响

虽然215.5 mm已经十分接近 $\alpha = 0.05$ 水平下 H_0 拒绝域的临界点(215.35 mm),但犯II型错误的概率仍然超过50%,这说明实验的设计需要完善。如果将样本容量增加,得到的结果将有明显改善。假如其余参数均不发生变化,只将样

本容量增加至200尾,此时根据公式(3)和公式(4)计算,当样本平均值 $\bar{x} = 218$ mm时,犯II型错误的概率将降至59.08%,检验效能为40.92%(图6);而当样本平均值 $\bar{x} = 215.5$ mm时,犯II型错误的概率将降至6.54%,检验效能达到93.45%(图6)。不同样本容量($n \in [50, 1000]$)条件下,计算得到的检验效能值的变化趋势(图7)。

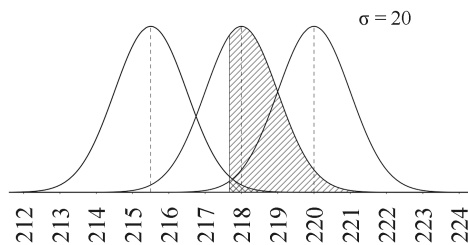


图6 与图1相比样本容量对检验效能($1-\beta$)的影响情况

平均值为220、218和215.5,标准差等于20,显著性水平 α 为0.05,样本含量等于200。217.67为平均值220的抽样分布的拒绝域临界点。斜线部分为平均值为218时犯II型错误的概率 β 的大小;网格线部分为平均值为215.5时犯II型错误的概率 β 的大小

Fig. 6 The effect of sample size on the statistical power ($1-\beta$) compared with Fig. 1

Means are 220, 218 and 215.5, standard deviation is 20, significant level is 0.05 and sample size is 200. 217.67 is the critical point of the rejection region for the sampling distribution with mean of 220. Oblique line area represents the probability of Type II error β when mean is 220, and grid line represents the probability of Type II error β when mean is 215.5

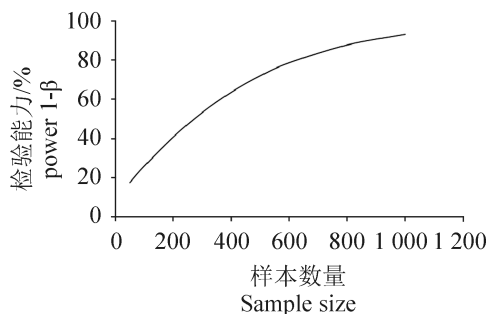


图7 不同样本容量得到的检验效能($1-\beta$)值

平均值为220和218,标准差等于20,显著性水平 α 为0.05

Fig. 7 Statistical power ($1-\beta$) values for different sample size

Means are 220 and 218, standard deviation is 20 and significant level is 0.05

4 总结与展望

II型错误是在水产科学研究中普遍存在,但却常常被忽视的问题。国外学者已经认识到该

问题的严重性,并开展了相关研究。而在国内这个问题至今尚未引起重视。不确定性是水产学研究的特点,水产科研工作者的主要任务之一即是尽量降低不确定性,减小由于误差给研究工作带来的误导和影响,从而给管理决策者提供更加全面可靠的建议。因此,在水产科学研究的各个领域,科研工作者都应该对II型错误分析加以重视,并在每一个相关工作中加以应用。在任何假设检验结果未能拒绝原假设时,均不要急于下结论或提出建议,而是计算检验效能或者 β ,并以图形呈现管理与决策者,便于他们更好地理解。

同时,由于管理与决策者的每个决定都关系到渔民的利益和资源的前景,因此管理决策者同样应该对II型错误问题有清晰的了解。他们要与科研工作者一起完成对未来的管理决策的设计工作,并要求研究人员提供的报告有明确的要求,对其中涉及到的II型错误问题都要有对 β 的计算^[6]。接受一个检验效能很低的结果甚至可能比什么都不做所带来的结果更糟糕,因为检验效能低很可能会误导管理和决策。因此,管理与决策者不能只关心研究结论,更要关注得到结论的过程,以避免危险的信号被乐观的表象所掩盖。

II型错误计算更多可用于实验设计,如确定低于某 β 水平时所需样本容量大小等^[12-15],来帮助科研工作者完善实验设计,服务于管理和决策。例如,在渔业资源评估工作中,常要面对的一个问题是判断在某一段时间内资源是否有明显变化。许多检验效能分析的结果显示,如果资源调查不充分(包括调查站位或航次不够),根据调查结果常常无法判断出资源水平下降的趋势,这样不能及时采取有效的管理措施,从而导致资源走向衰退^[16-26]。如果通过分析事先设定一个能够满足检验效能要求的调查站位(或航次)的最小值,则不仅可以提高资源保护和管理的效率,也能增加决策的可靠性。

需要强调的是,笔者旨在提醒科研工作者与管理决策者,在重视I型错误的同时也要重视II型错误,并非只强调II型错误重要而降低I型错误的重要性。事实上,两种错误的重要性要根据实际情况而定。只是在以往的显著性检验过程中,检验者更多强调I型错误,即认为要尽量避免把真实的情况否定掉。此时,检验者自己做

了一个无形的假设,即原假设都是重要的,而且一般是真实的。但事实上很多时候备择假设所描述的情况可能更重要,而且所做原假设也很有可能是错误的。因此,片面强调I型错误的重要性可能会导致判断错误,甚至引起严重的后果。现实中,水产科学研究的不同领域、不同实验问题对2类错误的要求会有所不同:有时I型错误重要些,有时II型错误更重要些,也有时2类错误同样重要。不应过分地强调检验的显著性水平而忽视检验效能。为了决定适当显著水平和应有的检验效能,不应设立绝对的标准^[10]。

今后,在水产科学研究中将会遇到很多诸如不能拒绝 H_0 和检验效能不高等问题,所以需要更多的II型错误分析以及更稳定、更高效的检验效能的计算方法,来不断完善水产科学的研究技术,促进水产科学健康发展。

参考文献:

- [1] Wang Y J, Liu Q, Fu D J. Summary of the applications of mathematical and statistical methods in fisheries research in China [J]. Transactions of Oceanology and Limnology, 2004(3): 69–73.
- [2] 明道绪. 生物统计附试验设计. 第4版. 北京: 中国农业出版社, 2008: 60–62.
Ming D X. Biostatistics and experiment design [M]. 4th ed. Beijing: China Agriculture Press, 2008: 60–62 (in Chinese).
- [3] Tang P C. The power function of the analysis of variance tests with tables and illustrations of their use [J]. Statistical Research Memoirs, 1938, 2: 126–149.
- [4] Peterman R M, Routledge R D. Experimental management of *Oregon coho salmon (Oncorhynchus kisutch)*: Designing for yield of information [J]. Canadian Journal of Fisheries and Aquatic Sciences, 1983, 40(8): 1212–1223.
- [5] Nickelson T E. Influences of upwelling, ocean temperature, and smolt abundance on marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon Production Area [J]. Canadian Journal of Fisheries and Aquatic Sciences, 1986, 43(3): 527–535.
- [6] Peterman R M. Application of statistical power analysis to the *Oregon coho salmon (Oncorhynchus kisutch)* problem [J]. Canadian Journal of Fisheries and Aquatic Sciences, 1989, 46(7): 1183–1187.
- [7] Holt R S, Gerrodette T, Cologne J B. Research vessel survey design for monitoring dolphin abundance in the eastern tropical pacific [J]. Fishery Bulletin, 1987, 85(3): 435–446.
- [8] Peterman R M. Statistical power analysis can improve fisheries research and management [J]. Canadian Journal of Fisheries and Aquatic Sciences, 1990, 47(1): 2–15.
- [9] Wagner T, Irwin B J, Bence J R, et al. Detecting temporal trends in freshwater fisheries surveys: Statistical power and the important linkages between management questions and monitoring objectives [J]. Fisheries, 2013, 38(7): 309–319.
- [10] 郭春彦. 对目前教育与心理统计书中I型及II型错误的分析[J]. 北京师范大学学报(自然科学版), 1989, 10(2): 57–64.
Guo C Y. The analysis of type I error as well as type II error in existing books of statistics in education and psychology [J]. Journal of Beijing Normal University (Natural Science), 1989, 10(2): 57–64 (in Chinese).
- [11] 张学中. t检验后II型错误 β 的估算和实际意义. 军事医学科学院院刊, 1991, 15(2): 136–139.
Zhang X Z. Estimations of type II error beta after t-test and their significance [J]. Bulletin of the Academy of Military Medical Sciences, 1991, 15(2): 136–139 (in Chinese).
- [12] Sokal R R, Rohlf F J. Biometry: The principles and practice of statistics in biological research [M]. 3rd ed. New York: WH Freeman and Company, 1995: 1–896.
- [13] Wellek S. Testing statistical hypotheses of equivalence [M]. Boca Raton, FL: CRC Press, 2002: 1–431.
- [14] Adelman D E. Scientific activism and restraint: The interplay of statistics, judgment, and procedure in environmental law [J]. Notre Dame Law Review, 2004, 79(2): 497–583.
- [15] Brosi B J, Biber E G. Statistical inference, Type II error, and decision making under the US Endangered Species Act [J]. Frontiers in Ecology and the Environment, 2009, 7(9): 487–494.
- [16] Paulik G Anchovies, birds and fishermen in the Peru current [M]// Murdoch W W. Environment: Resources, pollution and society. Stamford: Sinauer rand Association. Inc., 1971: 156–185.
- [17] Regier H A, Hartman W L. Lake Erie's fish community: 150 years of cultural stresses [J]. Science, 1973,

- 180(4092): 1248–1255.
- [18] Cushing D H. A link between science and management in fisheries [J]. *Fishery Bulletin*, 1974, 72(4): 859–864.
- [19] Pawishr H, Maccall A D. Climate variation and exploitation in the Pacific mackerel fishery [R]. California Department of Fish Game Fish Bulletin, 1978: 1–110.
- [20] Saville A. Discussion and conclusions of the symposium on the biological basis of pelagic fish stock management [R]. *Rapports et Process-Verbaux Des Reunions. Conseil International pour l' Exploration de la Mer*, 1980: 513–517.
- [21] Thompson J D. Climate, upwelling and biological productivity: Some primary relationships [M]//Glantz M G, Thompson J D. *Resource management and environmental uncertainty: lessons from coastal upwelling fisheries*. New York: John Wiley, 1981.
- [22] Selgeby J H. Decline of lake herring (*Coregonus artedii*) in Lake Superior: An analysis of the Wisconsin herring fishery, 1936–1978 [J]. *Canadian Journal of Fisheries and Aquatic Sciences*, 1982, 39(4): 554–563.
- [23] Clark S H, Overholtz W J, Hennemuth R C. Review and assessment of the Georges Bank and Gulf of Maine haddock fishery [J]. *Journal of Northwest Atlantic Fishery Science*, 1982, 3(1): 1–27.
- [24] Beverton R J H. Science and decision-making in fisheries regulation [R]// Sharp G D, Csirke J. *Proceedings of the Expert Consultation to Examine Changes in Abundance and Species of Neritic Fish Resources*. San José, Costa Rica: FAO Fishery Report, 1981, 2: 919–936.
- [25] Butterworth D S. Assessment and management of pelagic stocks in the southern Benguela region [R]// Sharp G D, Cskke J. *Proceedings of the Expert Consultation to Examine Changes in Abundance and Species of Neritic Fish Resources*. San José, Costa Rica: FAO Fish. Rep. 1983: 329–436.
- [26] Allen K R, Kirkwood G P. *Marine mammals* [M]//Guland J A. *Fish Population Dynamics*. 2nd ed. New York: John Wiley and Sons, 1988: 251–269.

Neglected Type II errors in the researches of fisheries science

WANG Yingbin¹, YU Cungen^{1*}, CHEN Yong²

(1. *School of Fisheries, Zhejiang Ocean University, Zhoushan 316022, China;*

2. *School of Marine Sciences, University of Maine, Maine 04469, USA)*

Abstract: Test of statistical significance uses samples to infer the statistic property of population from which the samples are taken. Two types of error, Type I (i.e., reject a true null hypothesis) and Type II errors (i.e., accept a false null hypothesis) can be made in a test of significance. Significance tests are commonly used in fisheries to identify statistical differences. Of the 6 major domestic fisheries journals, significance test methods were used in 2235 research articles within the recent 5 years (from January 2011 to April 2015), which accounted for 58.4% of all the research articles that were published in these 6 journals during the corresponding period. However, attention was only paid to the Type I errors in all these studies, and the Type II error was totally neglected. However, Type II error is not unimportant. The probability of having Type II error is often higher than that of Type I error, and sometimes the consequences are more serious for committing Type II errors. Neglecting Type II errors in fisheries is likely to result in fisheries mismanagement, consequently leading to overexploitation of fishery resources. In this study, the analyses of the Type II errors in fisheries researches were reviewed, and the potential consequence and implication of committing Type II errors in domestic fisheries researches were discussed; and the calculation method of the probability of having Type II error β (or statistical power $1-\beta$) was described using fisheries examples. The relationships between β and several main influence factors were analyzed. We strongly suggest that Type II errors be considered and evaluated in the future fisheries research to reduce the likelihood of making serious mistakes in fisheries management.

Key words: fisheries; significance test; Type II error; statistical power and Type I error

Corresponding author: YU Cungen. E-mail: cgyu@zjou.edu.cn

Funding projects: National Natural Science Foundation of China (31270527); Public Welfare Technology Application Research Project of Zhejiang Province (2015C33094); Natural Science Foundation of Zhejiang Province, China (LY13D010005); Young Academic Leaders Climbing Program of Zhejiang Province (pd2013222)