**统计功效与效应大小**

**华中师范大学心理学院 刘华山**

**一、统计功效**（**检验功效，效力，**Power）

统计功效指某检验能够正确地拒绝一个错误的虚无假设的能力。用1-β表示。

或说：当总体实际上存在差异，应该拒绝虚无假设时，正确地拒绝虚无假设的概率，或不犯β错误的概率 。在实验设计中，统计功效反映了假设检验能够正确侦查到真实的处理效应的能力。

统计功效的大小取决于四个条件：

1.两总体差异。

2.显著性标准α。

显著性标准α越大，则β错误越小，从而统计功效1-β越大。

3.检验的方向：当两总体差异一定，对于同样的显著性标准α，单侧检验比双侧检验的统计功效要大。

4.样本容量。样本容量越大，样本平均数分布的标准误越小，分布曲线越瘦削，统计功效越大。

**二、效应量 (**Effect Size,ES )

效应量，反映处理效应大小的度量。其实，两样本平均数的差异就是一个效应量。效应量表示两个总体分布的重叠程度。ES越大，表示两总体重叠的程度越小，效应越明显。

**三、效应量检验的功能**

1.效应量有助于我们判断统计上显著差异是否有实际的意义。

2.有些效应量，如相关系数，点二列相关系数的平方，，可以反映自变量解释因变量变异的百分比。

3. 在同一个实验中,如果有几个自变量,可以根据效应量大小对自变量的重要性排序。

4.原分析的基础。在元分析中，将各个不同的相关研究进行概括分析的基础便是各个不同研究的效应量。

5. 效果量的计算还为改进研究设计、 提高检验能力提供了根据。

APA出版手册第五版要求报告差异检验结果时一般要报告SE值。

**四、效应量和统计功效**

前述检验功效与两总体差异（或说处理效应大小）、样本容量、显著性水平、检验的方向性四个因素有关。而两总体差异大小、两样本分布的重叠恰恰是与效应量有关的概念。可见，效应量和统计功效有关。统计功效受效应量的制约。在检验方向、样本容量、显著性水平固定的条件下，效应量与检验功效有对应关系。见下表。【独立样本】

表 在0.05水平下假设检验的功效

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 样本容量 | 效应大小 | | |
| 0.2 | 0.5 | 0.8 |
| 单尾 | 10 | 0.11 | 0.29 | 0.53 |
|  | 20 | 0.15 | 0.46 | 0.80 |
|  | 30 | 0.19 | 0.61 | 0.92 |
|  | 40 | 0.22 | 0.72 | 0.97 |
|  | 50 | 0.26 | 0.80 | 0.99 |
|  | 100 | 0.41 | 0.97 | 1.00 |
| 双尾 | 10 | 0.07 | 0.18 | 0.39 |
|  | 20 | 0.09 | 0.33 | 0.69 |
|  | 30 | 0.12 | 0.47 | 0.86 |
|  | 40 | 0.14 | 0.60 | 0.94 |
|  | 50 | 0.17 | 0.70 | 0.94 |
|  | 100 | 0.29 | 0.94 | 1.00 |

**五、独立样本t检验的效应大小**

 例？在大学一年级新生中选取10名双性化学生和20名非双性化学生，对他们施测自尊量表。10名双性化学生得分的平均数为离差平方和SS1=670；20名非双性化学生得分的平均数为离差平方和SS2=1010。问两组平均数有无差异？（设α=0.01）



而效应量



这说明由双性化与非双性化造成的差异还是较大的。

2.Cohen's d

⑴指标1





（2）指标2

*t-t*检验值



df-*t*检验自由度

Cohen(1988)定义*d* 效应大小标准（解释）

d=0.2 小

d=0.5 中

d=0.8 大

3.Glass’ estimator gˊ

****

****为处理组的平均数，为对照组平均数，为对照组标准差。

**4.**Hedges' ĝ

（1）指标1 ĝ



（2）指标2 g



分母根号内为两独立样本方差分析中的误差均方。

5. 当对两独立组平均数之差进行检验时，也可用点二列相关系数的平方作为效果量的指标。不过也可用作相关样本的效应量的指标。其标准为：

0. 010 (效果小) ；0. 059 ( 效果中)；0. 138 (效果大)。

的计算公式为：

用作独立样本时，；用作相关样本时，。

6.对于两独立样本的平均数差异的检验，也可以对之作方差分析，用输出的作效应量。

**六、相关样本t检验的效应量**



2.两相关样本的效应量

指标公式



本公式与前述两独立样本t检验的效应量公式。即Cohen's d 指标1

3.点二列相关系数的平方（同前）

**七、检验的效应量**

**（一）检验的效应量**

1. Φ系数

类别变量的相关系数Φ系数也代表了效应大小。Φ系数（Φ相关系数）的计算公式是（适用于两列二分变量间的相关）：



2. Cramer’s φ



显然Φ系数是Cramer’s φ的特例。后者就是Cramer系数，或称克拉默系数V.

**（二）效应大小的判定标准**

当dfmin=1时，Φ=0.10表示低的效应；Φ=0.30表示中等的效应；Φ=0.50表示高的效应；

当dfmin=2时，Φ=0.07表示低的效应；Φ=0.21表示中等的效应；Φ=0.35表示高的效应；

当dfmin=3时，Φ=0.06表示低的效应；Φ=0.17表示中等的效应；Φ=0.29表示高的效应.。

当自由度越高时，效应量判定标准越低。

**八、方差分析中的效应量**

**（一）含义**

方差分析中的效应量用以测量处理效应（主效应、交互效应、线性比较等）与自变量关系程度的指标，它们可以被看作是自变量与处理效应之间的相关系数，它的平方可以解释为因变量总变异中各种效应的解释比例。

**（二）方差分析常用的四种效应量**

①Eta squared



②partial Eta squared



③omega squared



④the Intraclass correlation ρΙ΄

和是对样本中自变量和因变量关联程度的估计，是一个描述统计量；而和跨级相关是总体自变量与因变量关联程度的度量，是一个参数。每一个都有一个对应的。在一般情况下只要计算就足够了。

**(三)四种效应量的计算**

的计算

1. 义

Eta squared is the proportion of the total variance that is attributed to an effect.It is calculated as the ratio of the effect variance (SSeffect) to the total variance。

②公式



2.偏的计算

公式：某一效应的偏等于该效应的平方和除以该效应平方和与误差平方和的和所得的商。

3. ω2 的计算

4.跨级相关（Intraclass correlation）****的计算

①内涵

The intraclass is correlation an estimate of the degree of association between the independent variable and the dependent variable in the population for a random effects model.

②公式



**九、的应用**

可以在两独立样本的t检验、单因素方差分析、多因素方差分析、秩和检验、克-瓦氏H检验、单因素多相关组弗里德曼卡方检验中作为效应量指标使用。

（一）用作两独立样本的t检验中的效应量

（二）用作单因素方差分析中的效应量

（三）用作两因素方差分析中的效应量

（四）用作秩和检验中的效应量

先选定一组，算出其等级和ΣR。

根据下式算出期望等级和

, 其中应是为所选组的人数，N是总人数。根据下式算出Z值

再根据Z求出

（五）用作单因素多独立组克-瓦氏H检验( Kruskai- Wallis H test)中的效应量

先求出检验统计量H,然后运用下列公式根据H求出。

本例中H=4.94

（六）用作单因素多相关组弗里德曼卡方检验中的效应量

先求出值，再依据下公式求出

**十一、逻辑斯蒂回归的效应量**

Odds ratio（比值比，发生比之比，简写OR）



适于二个变量均是二分变量的情况，它也是逻辑斯蒂回归的效应量。

比值比的单位与Cohen’d不一样，因此同样的取值不是等价的。

**十二、相关系数的效应量**

相关系数的效应量指标，就是相关系数本身。皮尔逊积差相关系数r及点二列相关系数这是用得应用最为广泛的效应量之一。

根据Cohen的规定(1988，1992），0.10～0.29是小的效应；0.30～0.49是中等效应；等于或大于0.50是大的效应量。

另一个刻划两变量间关系强度大小的是确定系数。

**十三、效应量大小的标准**

计算出效应量后，如何解释、评价效应量的大小呢？评价的标准是什么？有以下几种方法。

**（一）为效应量规定数值标准**

不同的效应量指标标准是不同的。

1.Cohen(1988)定义*d* 效应大小标准（解释）（两个独立样本的t检验）

d=0.2 小

d=0.5 中

d=0.8 大

2.作为效应量的相关系数的评价标准

根据Cohen的规定(1988，1992），相关系数0.10～0.29是小的效应；0.30～0.49是中等效应；等于或大于0.50是大的效应量。

3.卡方检验中效应量Φ系数的评价标准与自由度有关。（见后）

4.点二列相关系数平方的评价标准

0. 010 (效果小) ；0. 059 ( 效果中)；0. 138 (效果大)。

5.总体效应量的评价标准

效果量标准是: 解释变异量 6%以下者,显示变量间关系微弱;解释变异量在6%以上到16%以下者,显示变量间属中等关系;解释变异量在16%以上者,显示变量间关系强。（Cohen J. 1982,1988)

6. 单因素方差分析中Cohen’s f的评价标准

根据Cohen的建议，f小于0.1为小的效应；f在0.25左右为中等效应；f大于0.4属于大的效应。

**（二）用自变量解释因变量变异的百分比来评价**

，，积差相关系数的平方等相关系数类效应量都可解释为：自变量可以解释因变量变异的百分比。

**（三）用两总体重叠部分的百分比来评价**

在两样本平均数差异的检验中，效应量可以用两个总体分布的重叠程度来解释

附表 Cohen’s d与两个样本分布的不重叠部分百分比

|  |  |  |  |
| --- | --- | --- | --- |
| Cohen的标准 | 效应量 | Percentile Standing | 不重叠部分百分比（%） |
|  | 2.0 | 97.7 | 81.1 |
|  | 1.9 | 97.1 | 79.4 |
|  | 1.8 | 96.4 | 77.4 |
|  | 1.7 | 95.5 | 75.4 |
|  | 1.6 | 94.5 | 73.1 |
|  | 1.5 | 93.3 | 70.7 |
|  | 1.4 | 91.9 | 68.1 |
|  | 1.3 | 90.0 | 65.3 |
|  | 1.2 | 88 | 62.2 |
|  | 1.1 | 86 | 58.9 |
|  | 1.0 | 84 | 55.4 |
|  | 0.9 | 82 | 51.6 |
| 大 | 0.8 | 79 | 47.4 |
|  | 0.7 | 76 | 43.0 |
|  | 0.6 | 73 | 38.2 |
| 中等 | 0.5 | 69 | 33.0 |
|  | 0.4 | 66 | 27.4 |
|  | 0.3 | 62 | 21.3 |
| 小 | 0.2 | 58 | 14.7 |
|  | 0.1 | 54 | 7.7 |
|  | 0.0 | 50 | 0.0 |

**十四、效应量的种类**

**（一）标准化的和非标准化的**

⑴标准化效应测量（一般文献效应量常指此）

标准化的效应量例如，相关系数r、Cohen’s d、odds ratio、Hedges’s g、Glass’sΔ等。

标准化效应量用于研究变量单位对于读者没有内在涵义时；或者多个研究被进行合并研究以期作出普遍结论的时候，而这些研究中全部或部分用的是不同的测度或量表；

⑵非标准化效应测量

例如：组间均值原始数值之差；非标准化回归系数等。

如果测量单位在实际水平上具有意义时，则我们常偏于用非标准化的效应测量。

**（二）利用样本统计量计算的，和不利用样本统计量计算出的**

**（三）基于相关系数与基于标准化平均差的效应量**

**（四）不同检验方法所用的不同的效应量**