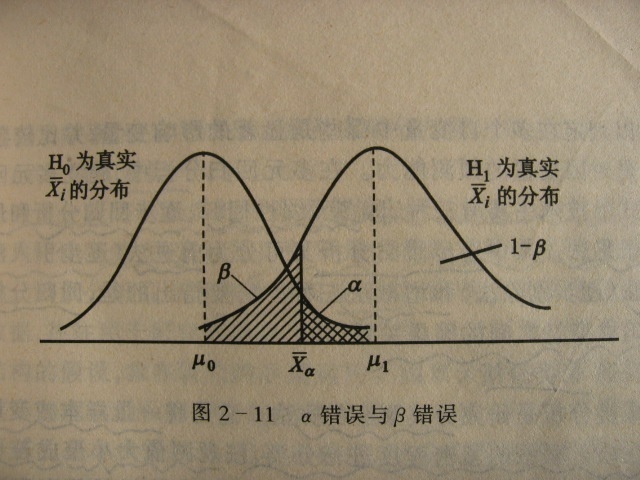
**统计功效与效应大小**

**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 。它表示某个检验探查出实际存在的差异，正确拒绝虚无假设的能力。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。当两总体实有差异越大，或处理效应越大，则假设检验的统计功效越大；

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；反之，α变小，1-β变小

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 (效应大小**，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。由计算出的ES大小，可由专门的表格中查出两样本分布的重叠的百分比。故效应量经常用两总体重叠的程度为指标，重叠的部分百分比越大，效应量越小。或以两个样本不重叠的程度为指标，不重叠的部分百分比越大，效应量越大。

**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

效应量检验，也就是要检验自变量作用的大小。它不同于差异显著性的检验。

统计显著性与实际显著性的区别：差异的统计显著性、相关的统计显著性只是告诉你在特定的条件下，这差异、这相关系数是存在的、并不是完全由抽样误差造成的，但并不意味着这差异有实际意义。大样本比较容易获得统计显著性的结果，但这并不意味着差异是有意义的。

2.有些效应量，主要是有相关意义的效应量，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把自变量的重要性排序。

4.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（的合成）。

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ES值。

美国心理学会1994 年发出通知,要求公开发表的研究报告包含效应量的测定结果。当具有统计上的显著性后,一定要计算效应量, 看你进行的研究是否有价值。

**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。可见，统计功效和效应量有关。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**

 上述效应量公式等价于



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）

已知 SS1=670；  SS2=1010，则



平均数差异的样本分布的标准误为



求效应量



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

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。此公式特点是用对照组方差代替合成方差。

本指标的使用范围同Cohen′s d指标。也有人将这里的效应量也称作d，其大小标准的判断也是0.2（小）,0.5（中）,0.8（大）。

**4.**Hedges' ĝ

（1）指标1 ĝ

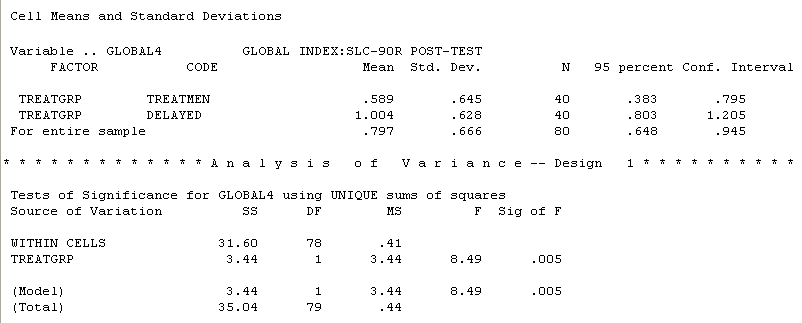


这只是对Cohen's d的小的修正。

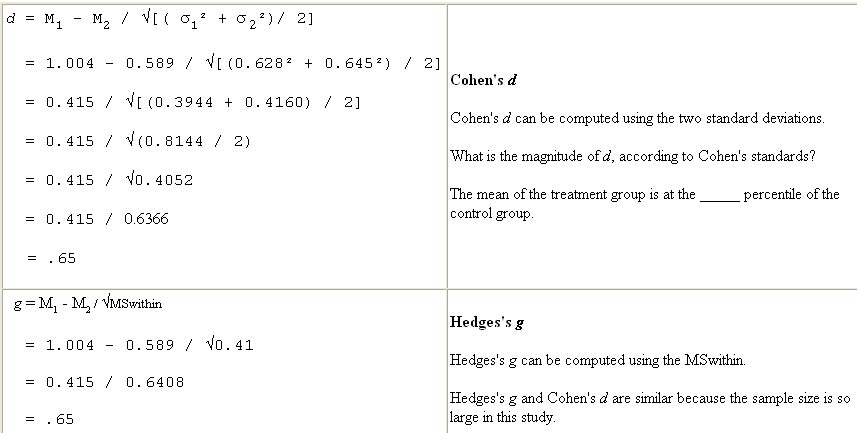
（2）指标2 g



分母根号内为两独立样本方差分析中的误差均方。



下表采用不同的两个公式计算效应量。前一种算法是以两样本方差的均值代替它们的联合方差。后一种算法是用方差分析组内方差的平方根作分母。本例求得的两个指标刚好相等。



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( point-biserial correlation coefficient ) 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、方差分析的效应量指标（见后）。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，。

此公式也可用于相关样本的t检验。本式显示标准化平均差与相关系数间的转化。

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（例见后）。

**六、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**

不同的效应量指标的评价标准是不同的。

1.Cohen(1988)定义*d* 效应量大小标准（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**

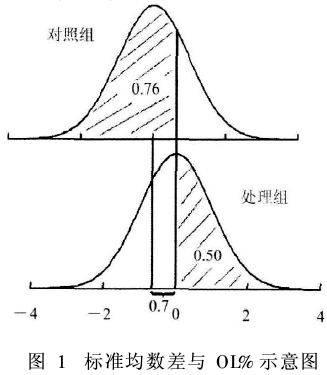
，，积差相关系数的平方等相关系数类效应量都可解释为：自变量可以解释因变量变异的百分比。

**（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。重叠部分比例越大，则效应量越小。

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 |



**七、相关样本t检验的效应量**



对照相关样本之差的检验的统计量，就知道SD的含义。



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）

**八、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（适用于两列多分类变量）



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）检验的效应量的评价标准**

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.。

**九、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）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

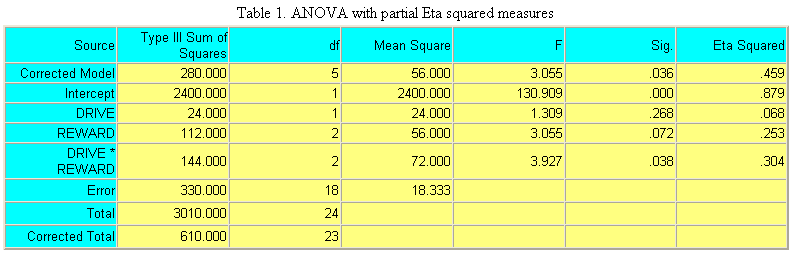


③omega squared

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中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。永远小于和。

**(三)四种效应量的计算**

**[](#760,36,幻灯片 36)**

该表是以drive和reward为自变量，以performance为因变量的二因素分析的结果.

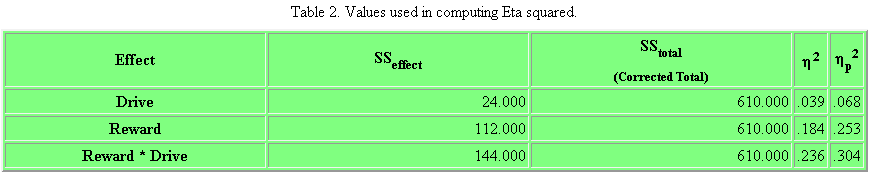
的计算

①含义

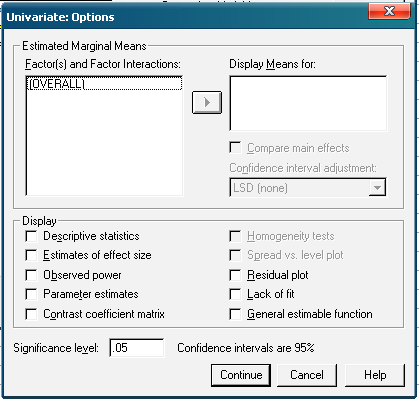
②公式



的计算（SPSS）

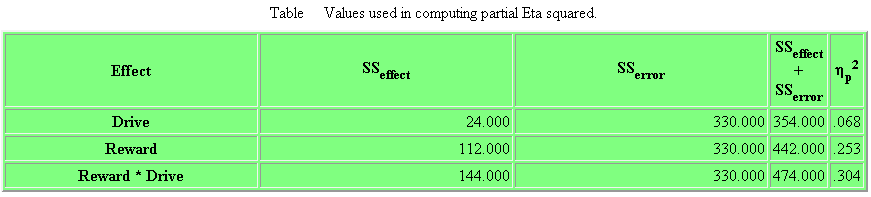


在SPSS软件的方差分析部分，在选择窗口“option”中可以选择“estimates effect size”即可输出和。

[](#758,32,幻灯片 32)

2.偏的计算

①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.Omega squared ω2 的计算

Omega squared is an estimate of the dependent variance accounted for by the independent variable in the population for a fixed effects model. The between-subjects, fixed effects。【用于固定效应模型】

**Note**

Do not use this formula for repeated measures designs。【不能用于重复测量的方差分析设计。】

1. 公式

关于效果量的高低判断方面, Cohen( 1982, 1988) 提出的标准是: 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. Statistical power and analysis for the behavioural sciences[ M] . Second dition , Hillsdalee, NJ: Erlbaum,1988.）

4.跨级相关（Intraclass correlation）****的计算

①内涵

Because it is for a random effects model it is not commonly used in psychology experiments. 【用于随机效应模型】

②公式



**十、单因素方差分析的效应量与统计功效**

**（一）效应量的计算**

f= 其中F为样本F统计量的观测值，n为每组样本容量。

这个f称为Cohen’s f

**（二）效应大小的判断**

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。

例：一项实验欲研究阅读时间长短对儿童阅读能力的影响。将儿童随机分配到3种阅读条件下，第一组阅读时间为5分钟，第二组为15分钟，第3组为30分钟。两周后测量儿童的阅读能力，得分如下。

表 阅读测试得分表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 第一组 | 第二组 | 第三组 | （K=3） |
|  | 10 | 15 | 10 |  |
|  | 14 | 20 | 12 |  |
| （n=5） | 12 | 17 | 6 |  |
|  | 8 | 8 | 12 |  |
|  | 11 | 15 | 10 |  |
|  | 11 | 15 | 10 |  |

方差分析结果为，所以三组阅读测验平均分数无显著差异。

计算效应量：f=,

可见，尽管方差分析的F值不显著，效应分析却显示大的效应。

**十一、回归或单因素方差分析中的效应量**

1.Cohen's **



为确定系数。

Cohen, 1988提出效应量大小标准。*≤0.02,*小，0.02<*<0.15*中*，>0.3,*大。

**十二、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

此时仍需作方差分析，而不是作t检验。

两种识记方法对回忆效果的影响的实验结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source | SS | df | Mean Square | F | p | Eta Squared |
| 识记方法 | 22.500 | 1 | 22.500 | 9.000 | 0.017 | 0. 529 |
| 误差 | 20.000 | 8 | 2. 500 |  |  |  |
| 总和 | 42.500 | 9 |  |  |  |  |

说明识记方法能解释回忆成绩变异量的53%。（指平方和的比例）

（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（两独立样本的非参数检验）

秩和检验中效应量的计算

设有两个独立组，每组5个被试。分别测量试验组对绿色信号灯的反应时，和对照组对蓝色信号灯的反应时。结果如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 试验组 | | 对照组 | |
| 反应时 | 等级 | 反应时 | 等级 |
| 539 | 2 | 759 | 7 |
| 480 | 1 | 890 | 8 |
| 600 | 5 | 1105 | 10 |
| 590 | 3 | 595 | 4 |
| 605 | 6 | 941 | 9 |
| ΣR | 17 |  | 38 |

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（较小的样本容量），N是总人数。根据下式算出Z值

再根据Z求出

本例中求得ΣR=17,，Z=-2.19,，说明信号颜色能解释反应时变异的53%。

**十三、逻辑斯蒂回归的效应量**

Odds ratio（比值比，发生比之比，简写OR）



在逻辑斯蒂回归中，称为发生比。

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。（见“六”中“二分变量的两独立组比较”）

**十四、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。

**十五、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。

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。

**（二）利用样本统计量计算的，和不利用样本统计量计算出的**

1.如果研究报告中没有报告统计量

例如当实验组与控制组的标准差相差不大时，有公式



2.如果研究报告中出现了统计量，则需视统计量的具体情况计算效应量。

（1）独立样本，研究结果以t表示时，其效应量：



（2）研究结果有表示时，其效应量为



**（三）基于相关系数的效应量与基于标准化平均差的效应量**

在相关设计中，一般采用相关系数作为效应量的指标；在实验组控制组设计中，一般采用标准化的平均差作为效应量的指标。二者可由点二列相关而得到转化。即t值可转化为点二列相关系数。

**（四）不同检验方法所用的不同的效应量**

**（五）元分析理论的不同学者提出的不同的效应量**