

等效检验及其软件实现

多数情况下的，我们作假设检验是想说明不同处理组间是有差别的，即所谓的优势研究(superiority study)，如t检验。而有时我们的目的可能是想说明两个方法是等效的(没有差别)，此时，决不能因为 $P > \alpha$ 就认为组间没有差别，它只是说明在统计上拒绝 H_0 的证据不足而已。

当我们想以较大的把握($1-\beta$)估计两组效果的不同并没有超过一个我们事先确定的界限(threshold)时，需用等效研究(equivalence study)。那么什么时候考虑用等效研究而不用优势研究呢？一些人为了说明一种方法是有效的，总是设计为一个对照组，一个处理组。但是，有些情况下，这样做是不道德的。比如，尽管一个常规的治疗方法(对照组)已经被证明是有效的，但为了评价一种新方法是否有效，就用这种新方法治疗一组病人，用安慰剂治疗另一组病人，而不是用常规疗法。

等效研究用于检验两个过程是同样的结果、新的方法同以往的标准方法相同，还可用于说明一段时间后产品的稳定性、两种服药途径等效否。该检验最早在药学方面开始广泛应用。

对于等效研究，可考虑用等效检验(equivalence test)。Equiv TestTM2.0软件能进行单侧和双侧的等效检验，包括两组、配对、 2×2 交叉设计的均数比较以及两组的率的比较。单侧等效检验用于说明新产品或药品同以往产品至少一样好；双侧等效检验用于说明既不更好也不更差。

下面将以两独立样本均数比较的等效检验为例对等效检验的操作过程进行介绍[1][2]。

例：用含有不同剂量的某微量元素的饲料喂养大白鼠，其体质量(g)变化结果如下。可否认为高剂量组(high dose)同对照组(control)的体质量变化无差别？

control: 519.80, 535.70, 491.70, 508.70, 565.50, 485.80, 520.70, 507.50, 428.90, 466.00

high dose: 465.00, 492.20, 466.30, 555.50, 528.10, 483.80, 502.80, 468.10, 519.20, 500.20

(1) 操作过程

① 数据格式

结果变量：“weight”。

分组变量：“dose”变量中的“1”代表control，“2”代表high-dose。

② 一般过程

从菜单选择

Analyze

Equivalence Testing

Means Independent Design...

弹出Specify Equivalence Test on Means...对话框(图1)。

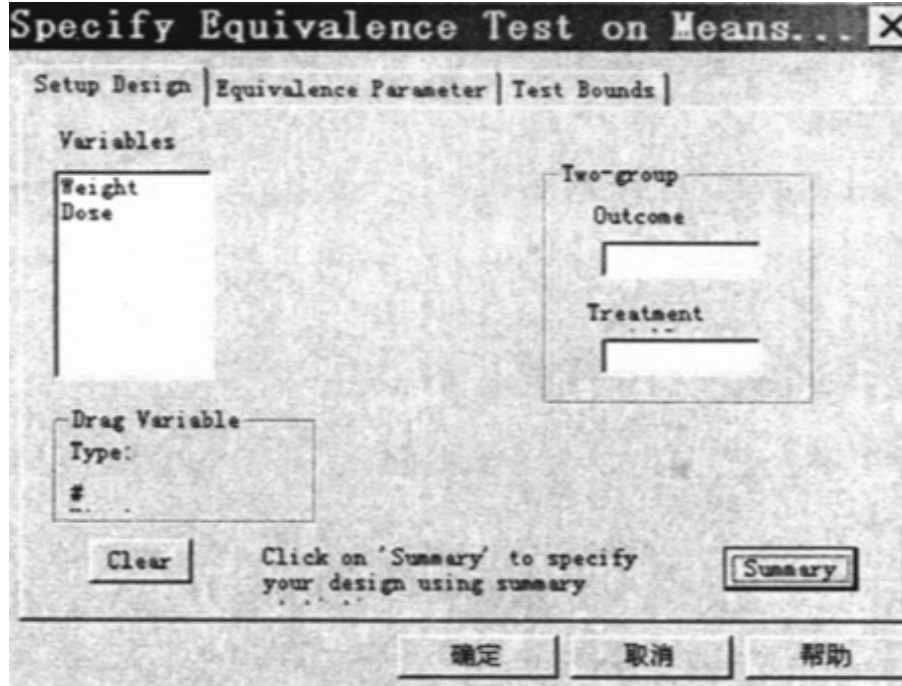


图1 两独立样本等效检验设计对话框

Fig.1 Designing dialogue box of equivalence test for two independent samples

将结果变量“weight”拖入“Outcome”框中；然后将分组变量“Dose”拖入“Treatment”中。

随即出现Set Treatment Order对话框（图2），由用户指定数据文件“dose”变量中的“Control”和High-dose“哪一个是Test（检验组）”，哪一个是Reference（参考组）。本例以“Control”作为参考组，以“High-dose”作为检验组。点击Equivalence Parameter出现如下对话框（图3）。

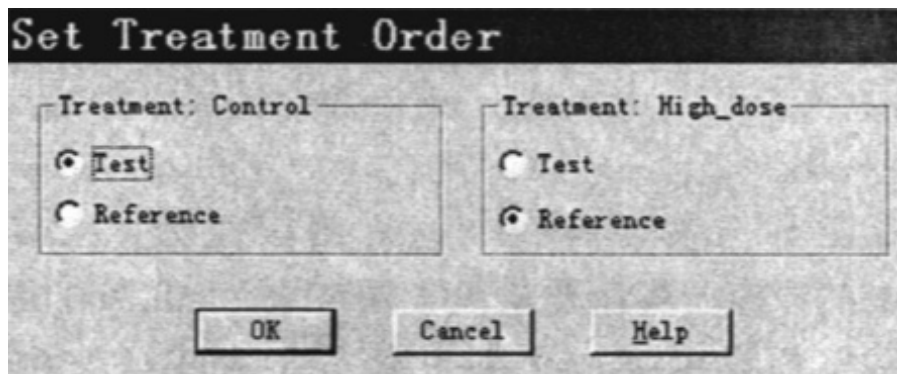


图2 设定组别对话框

Fig.2 Dialogue box for group setting

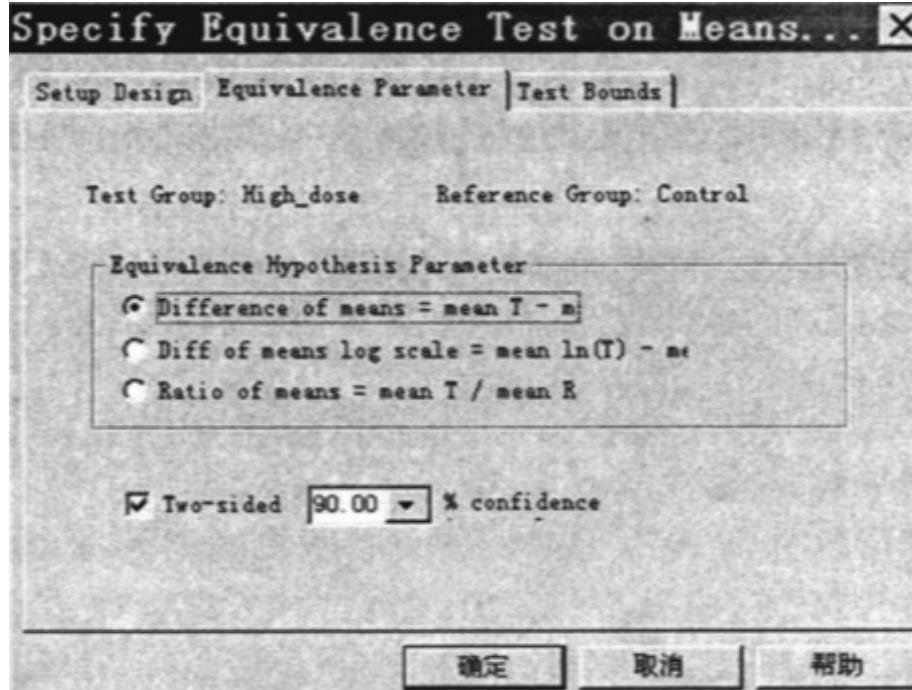


图3 两独立样本等效检验参数对话框
Fig.3 Dialogue box of the equivalence parameters

图中间的三个选项分别对应于

- Difference of means $\mu_T - \mu_R$
- Difference of means for log-scale $\ln\mu_T - \ln\mu_R$
- Ratio of means μ_T / μ_R

当假定资料为正态分布时，对于绝对数，选择Difference of means，对于相对数，选择Ratio of means。当假定资料为对数正态分布时，选择Difference of means log-scale。

当不知为何种分布时，对于绝对数，选择Difference of means，对于相对数，选择Difference of means log-scale。

本对话框还可以选择是否显示两组总体均数之差的置信区间，默认90%的置信区间。

本例选择 Difference of means 。

单击Test Bounds出现如下对话框(图4)。该对话框必须进行设置，否则将显示出错信息。

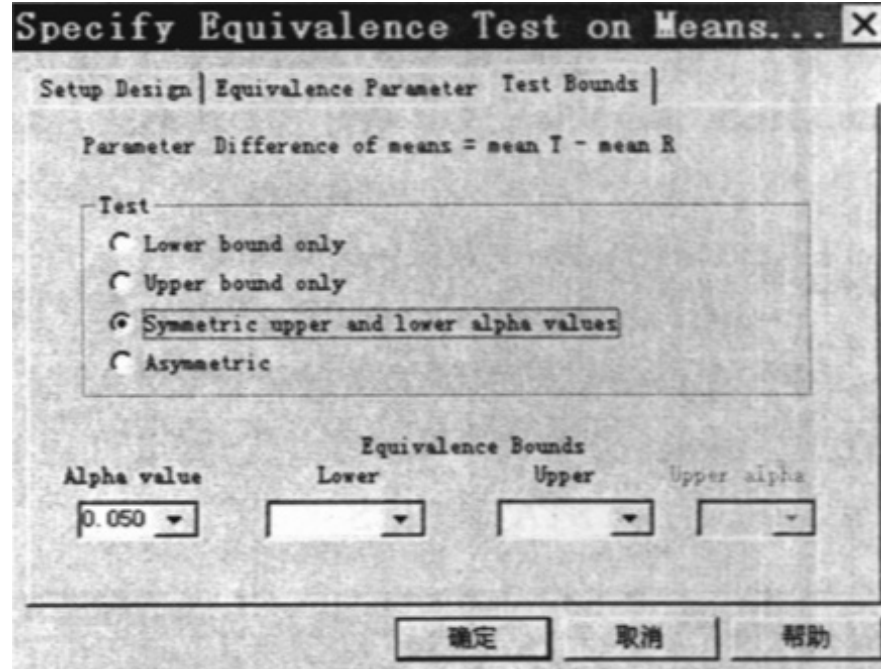


图4 两独立样本等效检验界值设定对话框
Fig4. Dialogue box for setting the test bounds

- Lower bound only: 单侧检验, 只检验下限
- Upper bound only: 单侧检验, 只检验上限
- Symmetric upper and lower alpha values: 对双侧对称的上下限进行检验
- Asymmetric: 对双侧非对称的上下限进行检验

根据上述不同的选项, 图中下侧的四个对话框会相应的被激活。系统默认的双侧显著性水准是0.05, 单侧为0.025。

绝对数的双侧检验界限可表示为 $\theta_L < \mu_T - \mu_R < \theta_U$

相对数的双侧检验界限可表示为 $(\delta_L - 1)\mu_R < \mu_T - \mu_R < (\delta_U - 1)\mu_R$

美国和欧洲的标准不同 (Difference of means log-scale 和 Difference of ratio of means)

美国FDA的标准是 $\delta_L = 0.8$ 和 $\delta_U = 1.25$

欧洲的标准是 $\delta_L = 0.7$ 和 $\delta_U = 1.43$

点击Lower 和(或)Upper, 根据所要进行的检验, 在下拉列表中, 默认的FDA和欧洲检验界限值将以均数的倍数形式显现出来。用户可选择需要的界限, 或输入新的值。这些值将作为所要进行的等效检验的上下界限。本例选择 ± 0.2 倍的均数作为上下界限。

(2) 结果输出

默认的结果输出为该检验过程的有关设定 (SPECIFICATIONS)、描述性统计量 (DESCRIPTIVE STATISTICS)、等效检验结果 (PARAMETRIC METHODS)。

① SPECIFICATIONS

Date: Monday, June 09, 2003 at 21:11:59
 Data Set: TOX
 Analysis: Means - Independent Group Design
 Outcome: Weight
 Grouping: Dose [Test: High_dose]

Equivalence Parameter: Difference of Means

90.00% CI: [-31.2843, 21.4643]

Equivalence Bound(s) [Lower]: -100.6060

[Upper]: 100.6060

Alpha Value(s) [Lower]: 0.0500
 [Upper]: 0.0500

说明：此部分结果主要给出进行该检验前的有关设定。

② DESCRIPTIVE STATISTICS

	Test	Reference
Mean	498.1199	503.0300
Standard Error (Mean)	9.3764	11.9755
Geometric Mean	497.3405	501.7155
Median	496.2000	508.1000
Standard Deviation	29.6509	37.8698
Variance	879.1795	1434.1270
Min	465.0000	428.8999
Max	555.5000	565.5000
Range	90.5000	136.6000
<i>n</i>	10	10

③ PARAMETRIC METHODS

Classical (shortest) Confidence Interval:

Difference=Mean for Test-Mean for Reference=-4.9100 Chow SC, Liu JP. (1992): Design and Analysis of Bioavailability and Bioequivalence Studies. Published by Marcel Dekker Inc., New York.

Confidence Bounds			
	Specified	Observed	Within equivalence limits?
Lower [5.00]% Conf. limit	-100.6060	-31.2843	Yes
Upper [5.00]% Conf. limit	100.6060	21.4643	Yes

Schuirmann OST/TOST:

Null Hypothesis L: Mean T- Mean R <= Lower Bound = -100.61

Null Hypothesis U: Mean T- Mean R >= Upper Bound = 100.61

t-Value

One-sided P -value to reject non-equivalence

	Specified	Observed	Specified	Observed
Null hypothesis L t-statistic	1.7340	6.2918	0.0500	3.117E-06
Null hypothesis U t-statistic	-1.7340	-6.9374	0.0500	8.757E-07

Both one-sided tests reject the null hypothesis, so equivalence within the specified equivalence bounds can be claimed.

Chow SC, Liu JP. (1992): Design and analysis of bioavailability and bioequivalence studies. Published by Marcel Dekker Inc., New York.

说明：该资料90%的可信区间为[-31.2843, 21.4643]，落在所设定的上下限[-100.6060, 100.6060]内；零假设的下、上限P值分别为3.117E-06和8.757E-07，均小于0.05，所以拒绝 H_0 ，接受 H_1 ，可认为两组结果等效。

若对当前检验过程的设定不满意，可在结果输出窗口的Rerun主菜单中进行重新设定。

配对设计、 2×2 交叉设计的均数比较以及两组率的比较的等效检验同上述过程基本类似，此不赘述。

参考文献：

[1] Rosner B. Fundamentals of biostatistics[M]. 5 th Ed. Harvard: Harvard University, 2000. 638-40.

[2] EquivTestTM software for the statistical analysis of equivalence and bioavailability studies[M]. Cork: Statistical Solutions Ltd., 2001.

参考文献：

[1] Rosner B. Fundamentals of biostatistics[M]. 5 th Ed. Harvard: Harvard University, 2000. 638-40.

[2] EquivTestTM software for the statistical analysis of equivalence and bioavailability studies[M]. Cork: Statistical Solutions Ltd., 2001.