

独立变量相依变量的区别？

DAND：心理学现象/统计学

vipqiaojin 2017-06-01 14:07:03 UTC #1

学渣求指导：

一、t检验分类？

是不是：单变量和多变量，多变量分为独立变量和相依变量？

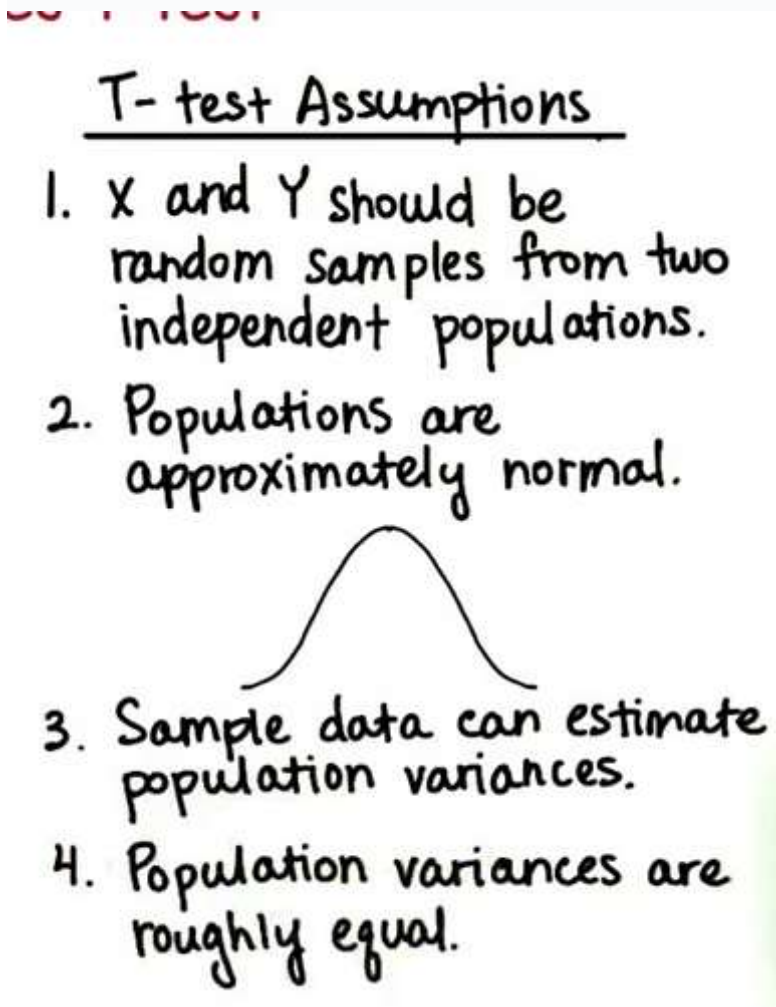
那么对于T检验的注意事项，该怎么理解？

1.相依变量怎么理解是独立变量？

2.总体近似正态，只有样本怎么才能确定总体情况呢？

3.样本数据近似总体，同上，何以知道总体？

4.总体方差大概相等？什么意思？总体和它的样本近似一致吗？怎么来确定呢？



Assumptions for the t-test

- Bivariate independent variable (A, B groups)
- Continuous dependent variable
- Each observation of the dependent variable is *independent* of the other observations of the dependent variable (its probability distribution isn't affected by their values). Exception: For the paired t-test, we only require that the pair-differences ($A_i - B_i$) be independent from each other (across i). [Note: "independent" and "dependent" are used in two different senses here. Just think of a "dependent variable" as one thing, and "observations that are dependent" as another thing.]
- Dependent variable has a *normal distribution*, with the same variance, σ^2 , in each group (as though the distribution for group A were merely shifted over to become the distribution for group B, without changing shape):

二、何时使用点估值？怎么理解？

三、cohen's的意义是什么？距离均值相差几个标准差，有什么意义？

四、项目检验为相依变量，怎么理解？

andy.li 2017-06-01 21:24:46 UTC #2

Hi @vipqiaojin ,

同学你的问题很犀利哦！👍

一、t检验分类

对的，大致可以分成“单样本和双样本 t 检验，
双样本又分为独立样本和相依样本”。

(请对 "自变量/因变量" 和 "独立样本/相依样本" 进行区分)

1) 既然“分为独立样本和相依样本”，那么就是不同类哦。

因此相依样本 **不能** 理解是独立样本...

这条是特指对 **独立样本** t 检验 的假设前提要求（注意图中红字），
不适用于 **相依样本** t 检验。

The image shows handwritten notes on a whiteboard. At the top, it says "Independent Samples t-test" in red. Below that, it lists "T- test Assumptions" with a list item: "1. X and Y should be random samples from two independent populations." To the left of this, there are two formulas: $\sum (x_i - \bar{x})^2$ and $\sum (y_i - \bar{y})^2$, with arrows pointing down to SS_x and SS_y respectively. Below these, it says $SS = SS_x + SS_y$.

2) 3) 所谓 assumptions，就是 **假设的** 前提条件，
最初有 **充足理由** 来假设即可，除非有反例能够推翻假设。

可以通过样本来推断：

从图形上可以使用 **Q-Q plot (quantile-quantile plot)** 来观察是否大致为正态分布。
不严格的话，使用直方图来看分布，大概是钟型图也是可以的...

4) 总体方差 (variances，注意复数) 大致相等，
是指 **x 和 y** 所代表总体的方差大致相等，
其实还是要通过样本来推断。

如果是相依样本，可以直接假设方差相等。

如果是独立样本，
简单的话，直接计算两组方差，看看是否大致相等...

严格的话，具体检验方法很多，可以使用 **F-test**、**Levene's test**、**Bartlett's test**、**Brown-Forsythe test**

至于后三个问题，大多是可以在 论坛 或 网络 中搜索到的。

请对自己的搜索能力善加培养，这对于数据分析师而言非常重要哟！

vipqiaojin 2017-06-02 14:38:49 UTC #3

感谢老师！

能否接着讲讲t检验 的前提？

Assumptions for the t-test

- Bivariate independent variable (A, B groups)
- Continuous dependent variable
- Each observation of the dependent variable is *independent* of the other observations of the dependent variable (its probability distribution isn't affected by their values). Exception: For the paired t-test, we only require that the pair-differences ($A_i - B_i$) be independent from each other (across i). [Note: "independent" and "dependent" are used in two different senses here. Just think of a "dependent variable" as one thing, and "observations that are dependent" as another thing.]
- Dependent variable has a *normal distribution*, with the same variance, σ^2 , in each group (as though the distribution for group A were merely shifted over to become the distribution for group B, without changing shape):

不知道我百度翻译的对不对.....

1.二元自变量。自变量为何出现俩？

2.因变量连续。什么维度的连续呢？时间？分组？

3.每个因变量相对独立。对于成对的数据，只需差值独立。是这样翻译理解么？

4.每组因变量服从正态分布。用样本来估计总体是否符合正态分布吗？样本小于30时发现非正态分布怎么做？大于30的样本不是正态分布怎么办？

上面的1.2.3.4 如果不成立，就不能使用t检验？

andy.li 2017-06-02 23:02:03 UTC #4

1) 这里其实是指：

自变量有两种取值（A/B、前/后、两种条件...），
不同取值对应两组因变量数据。

2) 连续和维度没有关系哦。

这里指的是 **变量种类**。

连续变量 最典型的例子是实数种类（身高、体重、时间...）

相对而言是 **离散变量**，其中包括：

- a. 分类变量（性别、国籍、公司...）
- b. 有序变量（排名、学历...）

3) 可以的。

第一句是对 独立样本检验 的要求，需要两组数据之间独立；

第二句是对 相依样本检验 的要求，仅需参与者之间互相独立。

4) 前面说过，只要有理由假设是正态分布其实就可以的，
除非用样本作为反例推翻前提假设。

样本如果小于 30，其实很难推断到底是不是正态分布。
不过 t 检验的优点在于，只要正态假设未被严重违反，
对结果的影响都不会很大...

大于 30 的话，
根据中心极限定理，抽样分布会趋近正态分布，
使用 z 检验也是可以的（当然会有偏差），
但使用 t 检验更加稳妥。

上面的 1.2.3.4 如果不成立，可以选择其他更加适合的检验种类，
当然这里没法列举，如果同学感兴趣，可以选购统计学教科书深入学习哦~

vipqiaojin 2017-06-04 01:57:00 UTC #5

第4点，有理由假设是正态分布即可。

怎么样才算正确的理由呢？

andy.li:

样本如果小于 30，其实很难推断到底是不是正态分布。
不过 t 检验的优点在于，只要正态假设未被严重违反，
对结果的影响都不会很大...

怎么样才算做严重违反？

只要不违反，那么全都可以假设当作正态分布吧？

andy.li 2017-06-04 02:18:26 UTC #6

前提假设有没有正当理由，其实放到具体情形，自然就清楚了哦~
空谈理论很难讲清楚，尽量举例说明吧：

通常一组相似人群的生理指标，
基本可以假设是呈正态分布的，
例如某个国某个年龄段人群的身高、体重等等。

如果整体数据来自 n 组差异比较大的人群，
那么很难假设呈正态分布，
比如让高中生和小学生同时参加难度较高的数学考试，
从整体来看，他们的考试成绩基本不会是正态分布
(更有可能是双峰分布)。

严重违反的话，虽然比较 silly，但请考虑扔骰子的例子：
1 到 6 投出的频数基本是 uniform 分布，
这种情况下基本不可能假设是正态分布的。

至于说“只要不违反，那么全都可以假设当作正态分布吧？”
其实很难回答呢...
只要能让你的读者接受你的前提，
不能随便找出反例应该就可以的。

既然说到前提假设了，不妨多说几句。
假设合不合理，大多是由是否符合常理，以及是否有反例决定的。

其实数学中的公设也是一种前提假设，
例如欧几里得《几何原本》中的平行公设，完全是“觉得”有道理，就决定接受了。
后来数学家们尝试，如果抛弃平行公设，去找反例会怎么样呢？
结果发现，居然可以推演各种整套理论，而且没有矛盾！
因此诞生了各种 **非欧几里得几何**。

有点跑题，请忽略这后半部分...

shine_zheng 2017-09-11 09:05:22 UTC #7

“3) 可以的。
第一句是对 独立样本检验 的要求，需要两组数据之间独立；
第二句是对 相依样本检验 的要求，仅需参与者之间互相独立。”
这里是不是写反了。。。

andy.li 2017-09-14 09:26:59 UTC #8

@shine_zheng

同学你好，

可以问一下你觉得这里写反了的理由嘛？😓

[shine_zheng](#) 2017-09-15 01:30:22 UTC #9

独立样本，被试间彼此独立；相依样本，比如同一组被试在前后两次实验，要求两波儿数据彼此独立。

所以我觉得应该改成：

“第一句是对 独立样本检验 的要求，需参与者之间互相独立；
第二句是对 相依样本检验 的要求，仅需要两组数据之间独立。”

[andy.li](#) 2017-09-15 01:45:11 UTC #10

展开来讲的话，其实存在两个层面的“独立”：

“波儿之间”与“参与者之间”。

独立样本：要求“波儿之间”与“参与者之间”同时独立。

相依样本：“波儿之间”不独立，仅需“参与者之间”独立。

也就是说，独立样本与相依样本的区别是在“波儿之间”的，
同学你再揣摩一下~

P.S.

换一种说法：

如果“相依样本，比如同一组被试在前后两次实验，
要求两波儿数据彼此独立”成立的话，

那么相依样本的“相依”体现在什么地方呢？

正是由于“相同”参与者进行两次测试，
两组数据才存在相依关系（不独立）的哟~

[首页](#) [分类](#) [FAQ/指引](#) [服务条款](#) [隐私政策](#)

采用 [Discourse](#)，启用 JavaScript 以获得最佳体验