

2001, 85-87, 193-203.

2. 三味工作室. SPSS V10.0 for Windows 实用基础教程. 北京: 希望电子出版社, 2001, 347.

3. Earl Babbie 著, 邱泽奇译. 社会研究方法(上). 北京: 华夏出版社,

2000, 182-187.

4. Earl Babbie 著, 邱泽奇译. 社会研究方法(下). 北京: 华夏出版社, 2000, 72-77.

5. 高惠肪等编译. SAS 系统 SAS/STAT 软件使用手册. 北京: 中国统计出版社, 1997, 389-393.

浅释 t 检验、方差分析的应用条件

兰州医学院卫生统计教研室(730000) 李娟生 丁建生 李曼生*

在作两个均数、多个均数的比较时,人们首先想到的是 t 检验、方差分析的 F 检验,也就是说这两种检验是作均数比较的常用方法,但因其为参数统计方法,故在应用时要注意其应用条件,一是正态性、二是方差齐性;为何要有这两个条件及在实际应用时该如何把握,我们作一浅释与探讨,使非统计专业人员在实际应用时能更好地把握,既能熟练运用,又能基本作到满足参数统计的应用条件。

一、两均数比较的 t 检验的条件

我们来看一下作 t 检验时计算的统计量:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (1)$$

它实际上是 t 分布变量的一个变通形式, t 分布的原形是

$$t = \frac{Z}{\sqrt{U/v}} \quad (2)$$

上式中分子 Z 是一个标准正态分布变量,分母中的 U 是一个服从自由度为 v 的 χ^2 分布变量,根据检验时的无效假设 $\mu_1 = \mu_2$,也可换成另一种形式 $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$,此时两样本均数之差的总体方差为 $\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$,由于实际上样本方差近似等于总体方差,在具体计算时可用 s_1^2, s_2^2 分别代替 σ_1^2, σ_2^2 ,设 s_1^2 的自由度为 $n_1 - 1, s_2^2$ 的自由度为 $n_2 - 1$,若两组的方差满足相等的条件,则两组数据可视为同一正态分布,用两者的加权平均来估计 σ^2 是合理的,即

$$\sigma^2 \approx s_c^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{\sum (x - \bar{x}_1)^2 + \sum (x - \bar{x}_2)^2}{n_1 + n_2} \quad (3)$$

同时不难证明, $(n_1 + n_2 - 2)s_c^2/\sigma^2$ 是一个服从 χ^2 分布的随机变量,自由度为 $(n_1 + n_2 - 2)$,而 $\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ 是一个

标准化正态分布的随机变量,因此据(2)式,则

$$\frac{\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}{\sqrt{\frac{(n_1 + n_2 - 2)s_c^2}{\sigma^2} / (n_1 + n_2 - 2)}} \quad (4)$$

是一个服从 t 分布的随机变量,将其约简后,便是(1)式的 t 检

验的统计量,因此,两均数比较的 t 检验要求数据正态性与方差齐性是参数统计量必须满足的条件。

二、方差分析

多个均数比较的参数统计方法是方差分析的 F 检验,而 F 统计量等于两个方差的比值:

$$F = \frac{s_1^2}{s_2^2} \quad (5)$$

实际上 F 分布的形式应为:

$$F = \frac{\xi_1/v_1}{\xi_2/v_2} \quad (6)$$

ξ_1, ξ_2 是服从 χ^2 分布的随机变量, v_1, v_2 为各自的自由度,从两个正态分布变量中随机取出的 n_1, n_2 个样本,样本方差分别记为 s_1^2, s_2^2 ,则 $(n_1 - 1)s_1^2/\sigma_1^2, (n_2 - 1)s_2^2/\sigma_2^2$ 是服从 χ^2 分布的随机变量,据式(6)有:

$$\frac{\frac{(n_1 - 1)s_1^2/\sigma_1^2}{n_1 - 1}}{\frac{(n_2 - 1)s_2^2/\sigma_2^2}{n_2 - 1}} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \quad (7)$$

服从 F 分布,上式若满足 $\sigma_1^2 = \sigma_2^2$,则(7)式就成了(5)式的形式。可见多个均数比较的方差分析同样要满足正态性与方差齐性。

那么在实际应用这些参数统计方法时,又该如何把握是否满足应用条件了呢?这里提供一种比较简便的方法,在应用 t 检验、方差分析时把握数据是否满足条件的一种粗略的估计办法:按惯例先计算各组的均数、标准差 $\bar{x}, s$,若 $\bar{x} \geq 2s$,且两组标准差的比值(大比小)不超过 2,则可视为满足条件,若 $\bar{x} \leq s$,则数据肯定不满足正态性,此时应对数据作适当变换后再作均数比较是合理的,同样若两组标准差的比值大于 2,则方差可能不齐,此时可用其方差比,去查 F 界值表(方差齐性检验用),以确认方差是否齐,若不齐,可改用 t 检验,当然对熟悉 SAS 软件或 SPSS 软件的非统计专业人员来说,问题可能就会变得简单一些。

参 考 文 献

1. Robert G. D. Steel, James H. Torrie. Principles and Procedures of Statistics. New York: McGraw-Hill Book Co., Inc. 1960.
2. 沈阳药学院主编. 高等数学—数理统计方法. 上海: 上海科学技术出版社, 1978.

* 兰州师范高等专科学校系计算机教研室