转 统计学---之样本方差与总体方差的区别

2018年01月24日 11:53:53 zxyhhjs2017 阅读数：29612

前段日子重新整理了一下这个问题的解答，跟大家分享一下，如果有什么错误的话希望大家能够提出来，我会及时改正的，话不多说进入正题：

首先，我们来看一下样本方差的计算公式：

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

刚开始接触这个公式的话可能会有一个疑问就是：为什么样本方差要除以（n-1）而不是除以n？为了解决这个疑惑，我们需要具备一点统计学的于总体、样本、期望（均值）、方差的定义以及统计估计量的评选标准。有了这些知识基础之后，我们会知道样本方差之所以要除以（n-1）是因为量才是关于总体方差的无偏估计量。这个公式是通过修正下面的方差计算公式而来的：

$$s^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

修正过程为：

$$s^2=\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)$$

我们看到的其实是修正后的结果：

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

对于这种修正的话是有相关的公式推导的。下面都会一一给出。

为了方便叙述，在这里说明好数学符号：

假设总体的均值为：μ

总体的方差为：σ^2

样本为随机向量：（x_1，$x_2 \cdots\cdots x_n$）

样本的均值为：$\bar{x}$

样本的方差为：s^2

前面说过样本方差之所以要除以（n-1）是因为这样的方差估计量才是关于总体方差的无偏估计量。在公式上来讲的话就是样本方差的估计量的期望方差。如下：

$$E(s^2)=\sigma^2$$

但是没有修正的方差公式，它的期望是不等于总体方差的

$$E(s^2)=E(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2)\neq\sigma^2$$

也就是说，样本方差估计量如果是用没有修正的方差公式来估计总计方差的话是有偏差的

下面 较好理解的公式推导过程：

$$
\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left((x_i-\mu)+(\mu-\bar{x})\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 - \frac{2}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)\left(\mu-\bar{x}\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\mu-\bar{x}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 - 2\left(\bar{x}-\mu\right)\left(\mu-\bar{x}\right) + \left(\mu-\bar{x}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 - \left(\mu-\bar{x}\right)^2 \le \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2
\end{aligned}
$$

也就是说，除非 $\bar{x}=\mu$ 否则一定会有

$$
s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 < \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2
$$

需要注意的是不等式右边的才是的对方差的“正确”估计，但是我们是不知道真正的总体均值是多少的，只能通过样本的均值来代替总体的均值。所 计量如果是用没有修正的方差公式来估计总计方差的话是会有偏差，是会低估了总体的样本方差的。为了能无偏差的估计总体方差，所以要对方差计算 正，修正公式如下：

$$
s^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)
$$

这种修正后的估计量将是总体方差的无偏估计量，下面将会给出这种修正的一个来源；

为了能搞懂这种修正是怎么来的，首先我们得有下面几个等式：

1. 方差计算公式：

$$
D(x) = E[x^2] - (E[x])^2
$$

2. 均值的均值、方差计算公式：

$$
E[\bar{x}] = \bar{x}\text{，}\ D[\bar{x}] = D[\frac{1}{n}\sum_{i=1}^{n}x_i] = \frac{1}{n}D[x_i] = \frac{1}{n}D[x](i=1,2\cdots,n)
$$

对于没有修正的方差计算公式我们有：

$$
\begin{aligned}
E(s^2) &= E(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2) \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - \frac{2}{n}\sum_{i=1}^{n}(x_i)(\bar{x}) + \frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}\right)^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - 2\left(\bar{x}\right)^2 + \left(\bar{x}\right)^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - \left(\bar{x}\right)^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(x_i)^2] - E[\left(\bar{x}\right)^2] \\
&= E[(x_i)^2] - E[\left(\bar{x}\right)^2] \\
&= D[(x_i)^2] + (E[x_i])^2 - (D[\left(\bar{x}\right)^2] + (E[\bar{x}])^2)
\end{aligned}
$$

因为：

$$
E[\bar{x}] = \bar{x} = E[x_i]\text{，}\ D[\bar{x}] = \frac{1}{n}D[x](i=1,2\cdots,n)
$$

丿 ：

$$
E(s^2) = D(x) - \frac{1}{n}D(x) = \frac{n-1}{n}D(x) = \frac{n-1}{n}\sigma^2
$$

在这里如果想修正的方差公式，让修正后的方差公式求出的方差的期望为总体方差的话就需要在没有修正的方差公式前面加上来进行修正，即：

$$\frac{n}{n-1}E(s^2)=\frac{n}{n-1}\times\frac{n-1}{n}D(x)=\frac{n}{n-1}\times\frac{n-1}{n}\sigma^2=\sigma^2$$

所以就会有这样的修正公式：

$$s^2=\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)$$

而我们看到的都是修正后的最终结果：

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

这就解释了为什么要对方差计算公式进行修正，且为什么要这样修正。

上面的解释如果有什么错误，或者有哪些解释不正确的地方欢迎大家指正。谢谢大家。希望能对大家有点帮助。

转载：http://blog.csdn.net/fuming2021118535/article/details/51290320