D 版的黄日华同学问了一个统计学基础课中一个灰常大众的问题，但是网上少有彻底的答案。为啥哩？因为要是能彻底理解这一问题，这个统计学基础课就学好了一半。话说要回答样本方差除以$\mathrm{n}-1$这个问题，我觉得还是要从头缕一缕比较好。

从头？啥是头？

总体和样本的概念。啥是总体？总体就是你要研究的对象的全体。很抽象？直观点说，比如，你想要研究你高三三班女同学的胸围问题，你滴总体就是她们的胸围。假设班上有 10 位女同学，刚好今天都在，那你就过去以 D 版的名义跟她们说我要研究一个课题，需要量一下你们的胸围。假如你是万人迷，10 位女同学都同意了。那你就获得了总体。假如有 4 个女同学不同意，你只量了 6 个，那么你获得的就是总体的一个样本。

假设我比较好事，请了三班女同学吃了顿饭，于是她们都同意给我量。然后我就获得了如下数据。(单位：cm)

X1=75，x2=75，x3=75，x4=75，x5=75，x6=85，x7=85，x8=85，x9=85，x10=85.

我马上就发现，总体期望值是 80.

于是我获得了**总体方差**$\boldsymbol{\frac{1}{10}\sum_{\mathrm{i}=1}^{10}(x_i-80)^2=25}$**。**

黄同学就不行了，只能量 6 个。他获得的那个样本如下。

x2=75，x3=75，x4=75，x5=75，x9=85，x10=85

他的样本期望值是 $\frac{1}{6}\sum_{i=2,3,4,5,9,10} x_i=78.33$

他的样本方差是$\frac{1}{5}\sum_{i=2,3,4,5,9,10}(x_i-78.33)^2=26.63$

问题来了，为啥求样本期望值是除以 6，而样本方差除以 5 呢？

这就涉及到统计学的估计量（estimator）的优良性了。通常，统计学对估计量的优良性有三个标准，无偏性（unbiasedness），一致性(consistence)，有效性(efficiency).

啥是估计量？就是根据你量到（或者目测？）的样本值，然后对获得的样本值进行运算，你会得到一个大概的值，这个值叫估计值。但是你肯定不会瞎算，对吧，你一般会想要有个法则，或者说公式，对了，这就是估计量。

比如要计算期望值，你会想到算术平均。可是黄同学对女同学 x5=85 比较有感觉，就想，我呀，要用加权平均，给我的女神加个权重。比如他的计算公式是

期望值 $=\frac{1}{7}\sum_{i=2,3,4,5,5,9,10} x_i$.

很聪明，居然还知道除以 7.

我在这里想问下同学们，这么做可以吗?

当然可以，(只要他没有四字真言)。

可是如果每个人都有自己的一套法则，这世界有点乱。

于是统计学家一讨论，还是要出台一些标准来比一比谁的法则好。这就有了以上说的三个标准。

在具体探讨这三个标准之前呢，我们先引入另外一个概念，就是，随机变量！

为啥要引入随机变量？因为各种估计量都是随机变量！

为啥？因为黄同学获得的是一个样本。如果换成另外一个同学去跟女同学们谈判的话，估计会多一两个女同学同意量。又如果黄同学换一天去，可能也会多一两个，当然也可能会少一两个。所以，一个估计量有可能有不同的值计算出来，这取决于谁去，什么时候去。

鉴别估计量优劣的标准主要围绕着估计量的随机性。随机性意味着估计量本身有自己的期望值（无偏性），概率极限（一致性），和方差（有效性）。

OK,啥是无偏。无偏就是说你的估计量，理论上，平均意义上要等于总体值。什么意思？意思就是，给你 1 次机会去获得 1 个样本，你根据你 1 个样本算出来的 1 个估计值，不会实际上等于总体值，但是这不重要，重要的是如果你有多次机会获得 n 个样本，你有 n 个估计值，你这 n 个估计值差不多要平均地落在总体值的两边。数学角度上讲，就是估计量的期望值要等于总体值。下面给个例子。

三班女同学胸围的总体期望值是一个固定的常数，用 X 指代。接下来，就比较玄了。这个常数呢，隐藏于 10 个女同学身上，或者说隐藏于柏拉图的理念世界中。由于黄同学不是温拿，无法获得准确值。他只能获得一个样本，5 个女童鞋的值。每个女童鞋的值可以被认为是 X 在这个现实世界中的实现，x_i, 但是它们每个也都是一个随机变量。然后我们可以认为它们互相独立，但是概率分布相同，所以它们都有一个共同的期望值，也就是$E(x_i)$ =X，虽然x_i基本上不可能刚好等于 X. 那么怎么利用这些 x_i 去估计 X 呢？我们用一个估计量，算术平均$\bar{x}$。也就是$\bar{x}=\frac{1}{n}\sum x_i$.　我

们说$\bar{x}$是 X 的一个无偏估计。这是因为 $E(\bar{x}) = E\frac{1}{n}\sum x_i = \frac{1}{n}E(\sum x_i) = \frac{1}{n}\sum E(x_i) = \frac{1}{n} * (n * X) = X$，虽然 $\bar{x}$ 也不会等于 X。黄同学也可以自己证明他自己的那个加权平均方法也是无偏的。（但是为什么统计学上不用那个来做估计量呢？因为那个加权平均不是有效估计。）但是如果他忘记应该除以 7，而是直接除以 6，那么就是有偏的。因为$E\frac{1}{6}(x_2 + x_3 + x_4 + x_5 + x_5 + x_9 + x_{10}) = \frac{1}{6}(7X) \neq X$.

所以我们看，在为总体期望值寻求一个无偏的估计量时候，我们要除以 n,而不是 n-1.

与此恰恰相反，在为总体方差寻求一个无偏的估计量时，我们选择了样本方差，我们是除以 n-1，而不是 n. 这就是答案了。具体证明见 33 楼。如果你看不懂那个证明，你需要把公式过一遍。我也帖在这里，顺便讲一下其中用到的公式。

$$E(S^2) = E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$

$$= \sigma^2$$

第二个等式运用到了期望值的线性，也就是 $E(aX + bY + c) = aE(X) + bE(Y) + c$. 连加符号也是线性。

第三个等式用到了$\sigma^2 = E(X^2) - \mu^2 \Longrightarrow E(X^2) = \sigma^2 + \mu^2$, $E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$是中心极限定理$Var(\overline{X}) = \frac{\sigma^2}{n} = E(\overline{X}^2) - [E(\overline{X})]^2$的变形。

对于这个问题，我现在发现，实际上要从数学上理解比较好，就是记住主要是无偏性的考虑。

如果一定要直观上理解，假设我们获得的样本有三个观测值，黑点，见下图

绿色点是样本期望值。

不难想象，假设这个样本是有代表性的，那么绿点是总体期望值很好的估计。这个点应该是出自于三个点的平均值，除以 3.

现在说方差。方差是用来描述随机变量离中心点的离散程度的。这个样本分布的离散程度主要地应该由绿点到两端黑点的距离来描述。如果除以 3，那中间黑点和绿点的距离也被赋予同样的权重，可是这个距离明显不具代表性。所以我们最好还是除以 2.

在小样本时，想象一下，至少有一个观测值离中心很近，对不对，可是这个离中心很近的距离不具代表性，所以算权重时，我们最好还是忽略它，所以我们除以 n-1。虽然平方的时候我们也平方这个距离，可是很小的距离的平方也相对较小，对不对，所以在连加的时候我们可以不忽略。

从另外一个意义上说，科学的态度是严谨，怎么严谨？就是保守一点，除以 n-1,让方差显得更大一点。

当然在大样本时，n 还是 n-1 不重要了，因为 1/n 和 1/(n-1)之间的区别就非常小了。

鉴于一致性和有效性相对较难，而基础课也一般不要求，所以就不讲了。