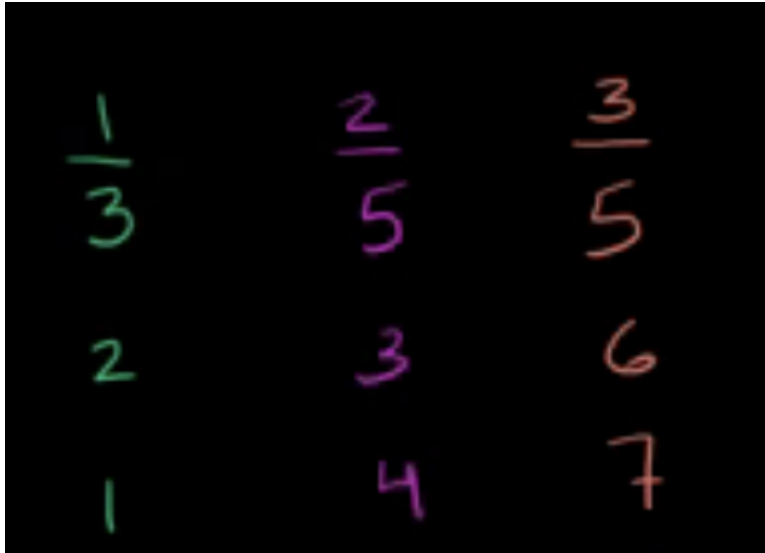


1. 总平方和 (sum of square total)

总平方和可以理解为计算方差时，不除以n的部分。

如下面的三组数据：



1	2	3
2	3	5
3	4	5
4	5	6
5	6	7

首先，计算总均值。总均值就是所有数据点的均值：

$$\bar{\bar{X}} = \frac{3 + 2 + 1 + 5 + 3 + 4 + 5 + 6 + 7}{9} = 4$$

总均值也可以看做是每组均值的均值：

$$\bar{X}_1 = \frac{3 + 2 + 1}{3} = 2$$

$$\bar{X}_2 = \frac{5 + 3 + 4}{3} = 4$$

$$\bar{X}_3 = \frac{5 + 6 + 7}{3} = 6$$

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3} = 4$$

总平方和：

$$SST = (3 - 4)^2 + (2 - 4)^2 + (1 - 4)^2 + (5 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + (7 - 4)^2 = 30$$

接下来，求自由度。自由度为总数据量减去1。这里的总数据量为9，则自由度为8。

2. 组内和组间平方和

这里我们想知道总平方和有多少是由于组内波动造成、多少是由于组间波动造成。

首先，求组内总波动，即组内平方和 (sum of squares within)。我们想知道的是总波动 (total variation) 有多少是因为各组内数据点同各自均值之间的差异造成的

$$SSW = (3 - 2)^2 + (2 - 2)^2 + (1 - 2)^2 + (5 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (5 - 6)^2 + (6 - 6)^2 + (7 - 6)^2 = 6$$

这表明，30的总波动中，有6个是来自组内波动。

组内自由度： $m(n-1)$ 。 m 表示组数； n 表示数据量。这个案例中自由度为6。

之后，我们求组间平方和(sum of squares between)。也就是说，有多少波动是来自组均值相对总均值的波动。

$$SSB = (2 - 4)^2 + (2 - 4)^2 + (2 - 4)^2 + (5 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (5 - 6)^2 + (6 - 6)^2 + (7 - 6)^2 = 24$$

每组内有多少数据就有几个平方差

组间自由度： $m-1$ 。 m 为组均值的数量。这里即为2。

这表明，30的总波动中，有24个是来自组内波动。

总结： 组内波动+组间波动=总波动； 组内自由度+组间自由度=总自由度