# 偏差

偏差又称为表观误差，是指个别测定值与测定的平均值之差，它可以用来衡量测定结果的精密度高低[1]。在统计学中，偏差可以用于两个不同的概念，即有偏采样与有偏估计。一个有偏采样是对总样本集非平等采样，而一个有偏估计则是指高估或低估要估计的量。

偏差不一定有害。尽管一个有偏采样会难以分析或引起不准确甚至错误的推断，但是有偏估计在某些情况下也有一些好的特性，例如较小的方差。

偏差分为绝对偏差和相对偏差、标准偏差和相对标准偏差来表示。

* 绝对偏差：是指某一次测量值与平均值的差异。
* 相对偏差：是指某一次测量的绝对偏差占平均值的百分比。
* 标准偏差：是指统计结果在某一个时段内误差上下波动的幅度。
* 平均偏差：是指单项测定值与平均值的偏差（取绝对值）之和，除以测定次数。
* 相对标准偏差：是指标准偏差占平均值的百分率。平均偏差和相对平均偏差都是正值。

Eg：分析铁矿石中铁的质量分数，得到如下数据：37.45，37.20，37.50，37.30，37.25（%），计算测结果的平均值、平均偏差、相对平均偏差、标准偏差。

解：平均值：(37.45+37.20+37.50+37.30+37.25)/5 = 37.34

各次测量的偏差分别是：0.11，-0.14，0.16，-0.04，-0.09

# 方差

方差（variance)是在概率论和统计方差衡量随机变量或一组数据时离散程度的度量。概率论中方差用来度量随机变量和其数学期望（即均值）之间的偏离程度。统计中的方差（样本方差）是各个数据分别与其平均数之差的平方的和的平均数。在许多实际问题中，研究方差即偏离程度有着重要意义。

方差是衡量源数据和期望值相差的度量值。

（1）

（2）设是常数，则

（3）设是随机变量，是常数，则有

（4）设与是两个随机变量，则

其中协方差

特别的，当是两个不相关的随机变量则

此性质可以推广到有限多个两两不相关的随机变量之和的情况。

（5）的充分必要条件是以概率1取常数，即

当且仅当取常数值时的概率为1时，。

注：不能得出恒等于常数，当是连续的时候，可以在任意有限个点取不等于常数c的值。

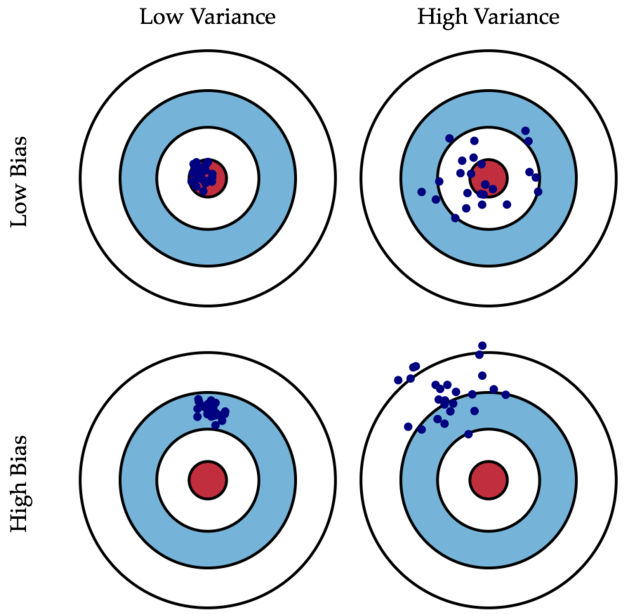
（6）

# bias & variance

偏差（bias）：描述的是预测值（估计值）的期望与真实值之间的差距。偏差越大，越偏离真实数据，如下图第二行所示。

方差（variance）：描述的是预测值的变化范围，离散程度，也就是离其期望值的距离。方差越大，数据的分布越分散，如下图右列所示。

注：预测值，即随机变量的取值；真实值，即随机变量的期望。



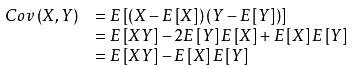
用误差衡量测量结果的准确度，用偏差衡量测量结果的精密度；误差是以真实值为标准，偏差是以多次测量结果的平均值为标准。

# 协方差

协方差（covariance或 covariation）分析是建立在方差分析和回归分析基础之上的一种统计分析方法。方差分析是从质量因子的角度探讨因素不同水平对实验指标影响的差异。一般说来，质量因子是可以人为控制的。回归分析是从数量因子的角度出发，通过建立回归方程来研究实验指标与一个（或几个）因子之间的数量关系。但大多数情况下，数量因子是不可以人为加以控制的。

在概率论和统计学中，协方差用于衡量两个变量的总体误差。而方差是协方差的一种特殊情况，即当两个变量是相同的情况。

期望值分别为与的两个实随机变量X与Y之间的协方差定义为：



从直观上来看，协方差表示的是两个变量总体误差的期望。协方差为0的两个随机变量称为是不相关的。

如果两个变量的变化趋势一致，也就是说如果其中一个大于自身的期望值时另外一个也大于自身的期望值，那么两个变量之间的协方差就是正值；如果两个变量的变化趋势相反，即其中一个变量大于自身的期望值时另外一个却小于自身的期望值，那么两个变量之间的协方差就是负值。

如果与是统计独立的，那么二者之间的协方差就是0，因为两个独立的随机变量满足。但是，反过来并不成立。即如果与的协方差为0，二者并不一定是统计独立的。

协方差的度量单位是的协方差乘以的协方差。而取决于协方差的相关性，是一个衡量线性独立的无量纲的数。

E代表期望，D代表方差，Cov代表协方差：

（1）

（2） （a, b是常数）

（3）

（4），

（5）

（6）

# 拉格朗日乘数法

在数学最优问题中，拉格朗日乘数法（以数学家约瑟夫·路易斯·拉格朗日命名）是一种寻找变量受一个或多个条件所限制的多元函数的极值的方法。这种方法将一个有n 个变量与k 个约束条件的最优化问题转换为一个有n + k个变量的方程组的极值问题，其变量不受任何约束。这种方法引入了一种新的标量未知数，即拉格朗日乘数：约束方程的梯度（gradient）的线性组合里每个向量的系数。此方法的证明牵涉到偏微分，全微分或链法，从而找到能让设出的隐函数的微分为零的未知数的值。

**拉格朗日乘子法的用途：**

从经济学的角度来看，代表当约束条件变动时，目标函数极值的变化。因为∂F/∂M=λ，当M增加或减少一个单位值时，F会相应变化。例如，假设目标函数代表一个工厂生产产品的数量，约束条件限制了生产中投入的原料和人力的总成本，我们求目标函数的极值，就是要求在成本一定的条件下，如何分配利用人力和原料，从而使得生产量达到最大。此时便代表，当成本条件改变时，工厂可达到的生产量最大值的变化率。

**具体方法：**

①假设需要求极值的目标函数 (objective function) 为，限制条件为=M

②设

③定义一个新函数

④则用偏导数方法列出方程：

⑤求出的值，代入即可得到目标函数的极值。

扩展为多个变量的式子为：，则求极值点的方程为：

（即为、……等自变量）

**实例：**

Eg：求此方程的最大值：，同时未知数满足。

解：因为只有一个未知数的限制条件，我们只需要用一个乘数。

设

则。

将所有方程的偏微分设为零，得到一个方程组，最大值是以下方程组的解中的一个：

# 参考

算法--偏差，方差，标准差，协方差，相关系数及相关理解：

<http://blog.csdn.net/sinat_26492471/article/details/52923545>

方差，标准差，协方差：

<http://www.cnblogs.com/cvlabs/archive/2010/03/26/1696978.html>

偏差和方差有什么区别？：

<https://www.zhihu.com/question/20448464>

拉格朗日乘子/拉格朗日乘数（Lagrange multiplier）：

<http://www.cnblogs.com/emanlee/archive/2012/02/06/2339510.html>

[Math & Algorithm] 拉格朗日乘数法：

<http://www.cnblogs.com/maybe2030/p/4946256.html>

百度文库：

<http://baike.baidu.com/link?url=ZK6PPGMZ54vIE8npVxhuxs09hlLRzEXg5Ugrldhl1oMinQZaeWY7qPi726Dn9-ADTl6NmXdoAyJbXZIb6dZjoiaT1LLDaiM3p_cG4hOrfRy>

<http://baike.baidu.com/link?url=bahFQ23ojRfyiBKm7AESLZ_lENW--fRYhUJTrk2je6hxpQKxiLigdR0t98xFM_FHGX6MSIZJscSfIt_sAGWhqWrxOMG04F0Ud6NIbbk-4-3>

<http://baike.baidu.com/link?url=M8OCQJB8q00WVKjyq33lS7aQNM4XSzdCdV_o8KeZ8HxeXLKJuWpwnQstlRUXIQmCfTySETPg7d05lI2Z3RFh7Ds5hgS-owRyYY3EGglVL1Q5-bG1svym8zJXh4-OhCZd>

<http://baike.baidu.com/link?url=bpigp2M9qWiSvqjHbilIE9qEQhMR2ZE7Ylr-66ND-7Eo6g5k5hHjCptjJisyaTmY_W9klOQRhmjpbcFDLFNcQkXElcalHWJPRyTrbv0YBh2w2dO1YQbhmFRMCFsFT5WJXsQsRjpvHJtxbzb4F5a6zLjBlW3u9ZFWx8aX_BMs01RnKka-4CVEIJNnugU4MbWrHD-OQIhwmcN9O1wCuIg5QHEVM2IJZfE4CpjQavAn6aUQdI0M-6lP1UoZsQstTWxcVTxUpZIYOoDfcpefS_1SXK>