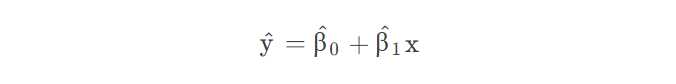
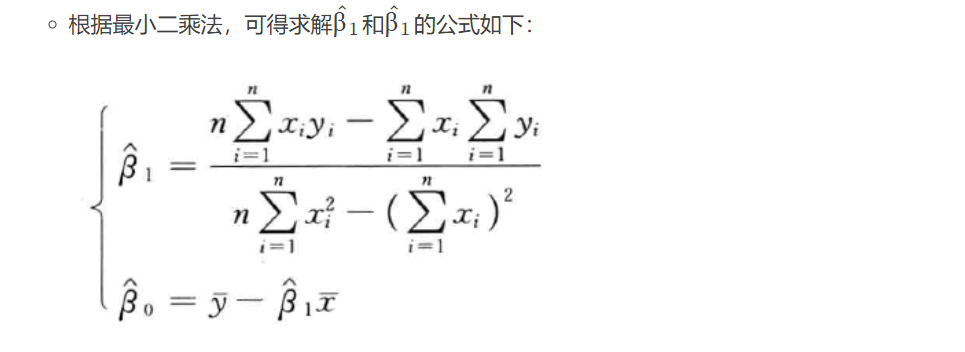
**问题一**：分别求得与三者的关系，再根据关系填补缺失值，计算年总水流量和年总排沙量。

1.数据预处理：首先合并六个表格，即将每一年的数据合并到一个表格中。

考虑到含沙量数据缺失值太多，不宜使用插值法处理缺失项，因此筛选出含沙量不为空白的行，将数据以8：2的比例划分为训练集和测试集，先以训练集来确定含沙量与时间、水位、水流量的关系。通过绘制散点图发现，含沙量与水位、水流量具有线性相关性，可以用线性回归方程分别刻画（严谨一点也可以用皮尔逊相关性系数定量计算）；而在每一年有一定的周期变化规律（夏秋季含沙量高），可以描述与时间的联系。

2.一元线性回归方程的求解：





设含沙量为y，水位为x1，水流量为x2，对数据进行线性回归处理得到：

y=b1x1+a1

y=b2x2+a2

通过合并两个一元回归方程，我们建立一个多元回归模型：

y=b1x1+b2x2+c

该方程刻画了含沙量与水位、水流量的关系。用该方程得到的数据与测试集进行误差分析，我们可以估计所得回归模型的准确度。

3.根据所得多元回归方程，我们可以计算出原数据集中含沙量的空白值，并将其填入表中。

记水流量为Q，含沙量为C，每段记录时间间隔为t，年总水流量为

年总排沙量为

**问题二**：

水沙通量指单位时间内通过某一断面或区域的水流和泥沙量的总和，记为T。则有：

*T=QC*

将每年水沙通量数据计算汇总排列，方便后续处理。

1.分析其突变性，可以采用Mann-Kendall 检验法，其基本思想是比较时间序列数据中每一对观测值，计算它们之间的顺序关系，从而评估数据的趋势是否显著。MK检验法参考网址：

<https://blog.csdn.net/aaakirito/article/details/116600294>

根据标准正态分布来判断 Z 值的显著性：

* 选择显著性水平（如 0.05 或 0.01），查找相应的 Z 临界值。
* 如果 |Z| 大于 Z 临界值，那么拒绝原假设，认为存在显著趋势。

2.分析其季节性，可以使用STL(季节性分解的局部回归)方法来分解时间序列数据，分解之后，可以将原始数据、趋势成分、季节成分和随机成分进行可视化。季节成分通常会表现为平滑的曲线，展示在特定周期内数据随时间变化的规律。

3.分析其周期性，可以通过谱分析判断其存在：对水沙通量数据进行离散傅里叶变换，然后绘制频谱图，若数据具有周期性，则频谱图上会在信号的主频率处出现显著的峰值；也可以将傅里叶变换结果的幅值平方绘制功率谱，在具有周期性的时间序列中，功率谱将显示出在主频和谐波处的明显能量集中。

**问题三**：采用lstm模型对未来2年的水沙通量进行预测。LSTM 对输入数据的值范围较为敏感，因此常见的做法是对数据进行归一化处理，使数据在 (0, 1) 之间，将时序数据转为 LSTM 模型需要的格式，使用过去的数值来预测未来的数值。对预测的数值进行反归一化，得到未来2年水沙通量数据，并以问题二中方法分析其季节性，将一年分为雨季、常规时期和稳定期，根据季节性来制定最优的采样监测方案：雨季按原频率每天检测1次，常规时期每三天检测1次，稳定期每周检测一次（具体检测频率可依季节时长而定）。

**问题四**：为简化模型，将一天测量的河底高程的数据取均值，视为该水文站的河底高程值。要预测实际效果和河底高程变化，首先要分析水沙通量与河底高程的相关性，先计算出对应时间的水沙通量均值，如2016年6月8日测量了河底高程，则找到2016年6月8日相应的水沙通量数据，计算其均值。将得到的对应几组水沙通量与河底高程数据进行回归分析，目的是通过水沙通量数据得到河底高程值。

每年的“调水调沙”在6-7 月进行，为分析“调水调沙”的具体效果，可以先以1~6月份的水沙通量数据进行模拟，利用lstm模型预测6-12月份数据，再计算真实值与预测值的差，得到调水调沙对水沙通量的实际影响效果。

如果不进行“调水调沙”，利用lstm模型预测10年后水沙通量（这里只能用2021年7月以后数据训练模型，因为使用之前的模型预测相当于进行了“调水调沙”），再利用水沙通量与河底高程的关系得到河底高程的值。

附件3真不知道怎么用，他最新的数据是2023年但前两个附件数据都只到2021年。问题四我做出来的回归效果感觉非常不好，但是没有别的思路了，你们看有什么妙招？