
高中组
个人项目
10 年级

黄浦江上游水质变化统计回归模型探讨

黄浦江上游水质变化统计回归模型探讨

摘要

本课题主要是根据作者从黄浦江水质考察过程中所获得的第一手资料，寻找合适的统计方法，然后利用统计软件建立统计回归模型，最后对模型进行了统计假设检验。按照本论文建立的模型，可以方便地对全区域的环境状况进行监测，比如可以根据模型对任意给定的区域水温，PH 值，化学需氧量进行估算。如果在将来的实际监测中发现和此模型预测的数据出入较大，就应该对该区域进行环境调查，是否有新的污染源出现，以及如何评估治理效果，是否到达目的。

在此文的撰写过程中，作者还对如何建立类似统计回归模型做了探索。所采用的方法完全可以应用在其它研究任何两个或两个以上变量有无（线性）相关关系，及相关关系如何的统计模型中。

需要指出的是，作者所利用的软件是一台图形计算器和 MICROSOFT OFFICE 中的数据分析包。读者可以利用其他类似的软件包，如 R, SSPS 其他工具进行运行数据，其输出结果应该一致。同时，作者对如何阅读解释 EXCEL 回归模型输出的报告，以及如何利用输出报告来改进模型做了探索。特别的，读者可以从本文了解如何把一个变量的高次线性回归(统计里的线性 linearity 通常是指因变量, 或叫 response variable 对自变量的系数是线性的) 转化为多变量的一次线性回归。类似的方法还可以通过取对数把指数或幂函数关系转化为线性关系，然后利用 EXCEL 处理转化为线性函数的回归问题。

全文分为四部分：

第一部分对数据的采集做了详细的描述。

第二部分对数据进行了简单的分析，然后详细讨论了如何寻找、建立、改进，最后获得比较理想的定量统计模型，同时对如何阅读回归模型检验报告（Summary output）做了详细说明。

第三部分对所建立的数学模型和现实情况进行比较，推测产生样本数据的可能原因，从环境科学的角度对模型的合理性进行定性验证。

第四部分总结了本文得出的主要结论：

1. 考察区域水温变化模型： $\text{水温} = -0.000969 \cdot \text{距离平方} + 0.174645 \cdot \text{距离} + 24.2174$
2. 考察区域溶解氧变化模型： $\text{溶解氧} = 8.22145 - 0.035348 \cdot \text{离开龙王山距离}$
3. 考察区域 PH 值、溶解氧具有弱相关性，二者都与周边生态环境及工业排放相关。
4. 探讨了如何将一元简单回归、一元高次多项式回归和多元线性回归及模型检

验方法应用于环境科学的研究中，取得了预期效果。

关键词：统计；分析； 回归模型；模型检验；相关系数；F 检验；T 检验；环境因果分析。

目录

第一部分：采集并分析水质数据.....	5
1.1 采集数据.....	5
1.1.1 采集河流断面位置及顺序.....	5
1.1.2 采集河流介绍.....	6
1.1.3 采集设备及方法.....	6
1.2 记录，整合数据.....	7
第二部分：数据的相关性分析及统计回归模型.....	11
2.1 各区域水温变化回归模型.....	11
2.2 改进模型.....	13
2.3 回归分析报告的统计学说明，模型的假设检验.....	15
2.4 对各个区域 PH 值和溶解氧的相关关系的探讨.....	16
2.5 平均溶解氧回归模型.....	17
第三部分：对统计模型的环境验证.....	19
3.1 分析并对比孝丰大桥及乌巷大桥河流断面数据.....	20
第四部分：结论.....	21
致谢.....	21
参考文献.....	22

第一部分：采集并分析水质数据

1.1 采集数据

1.1.1 采集河流断面位置及顺序

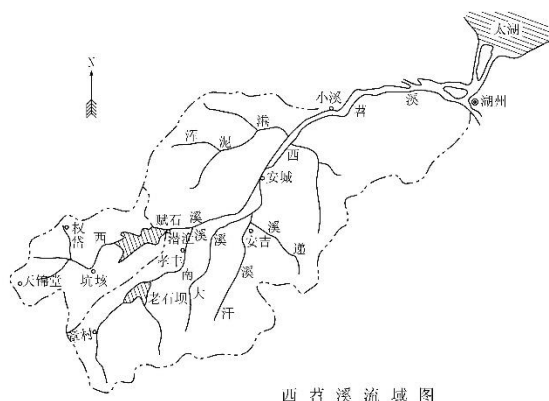
此次黄浦江水质考察来到了位于黄浦江上游的浙江省湖州市安吉县以及湖州市市区。最先采集的数据为黄浦江源——龙王山。随后顺着西苕溪河流往东北方向，顺次采集孝丰东门大桥、乌巷大桥、安城大桥、梅溪大桥、港口大桥和白雀大桥的河流断面数据。如下图所示。



1.1.2 采集河流介绍

西苕溪属于苕溪水系，是浙江八大水系之一。苕溪水系分为西苕溪以及东苕溪，两溪在湖州市白雀塘桥汇合经长兜港注入太湖。苕溪干流长 158 公里，流域面积 4576 平方公里，河道落差 779 米。

西苕溪，又名龙溪港，因在湖州城区以西，故名。上游有南溪、西溪两源，西溪为正源，源于浙江安吉和安徽宁国两县交界的天目山北侧南北龙山之间的天锦堂，山峰海拔 1415 米，东北流至安吉县塘浦乡汇合南溪后始称西苕溪。



西苕溪流域图

1.1.3 采集设备及方法

(1) 采集设备



采集设备主要分为两类，第一类为水体采样设备，第二类为数据采集设备。第一类设备用于采集水源，包括采样桶，采样瓶等（如左图所示）。

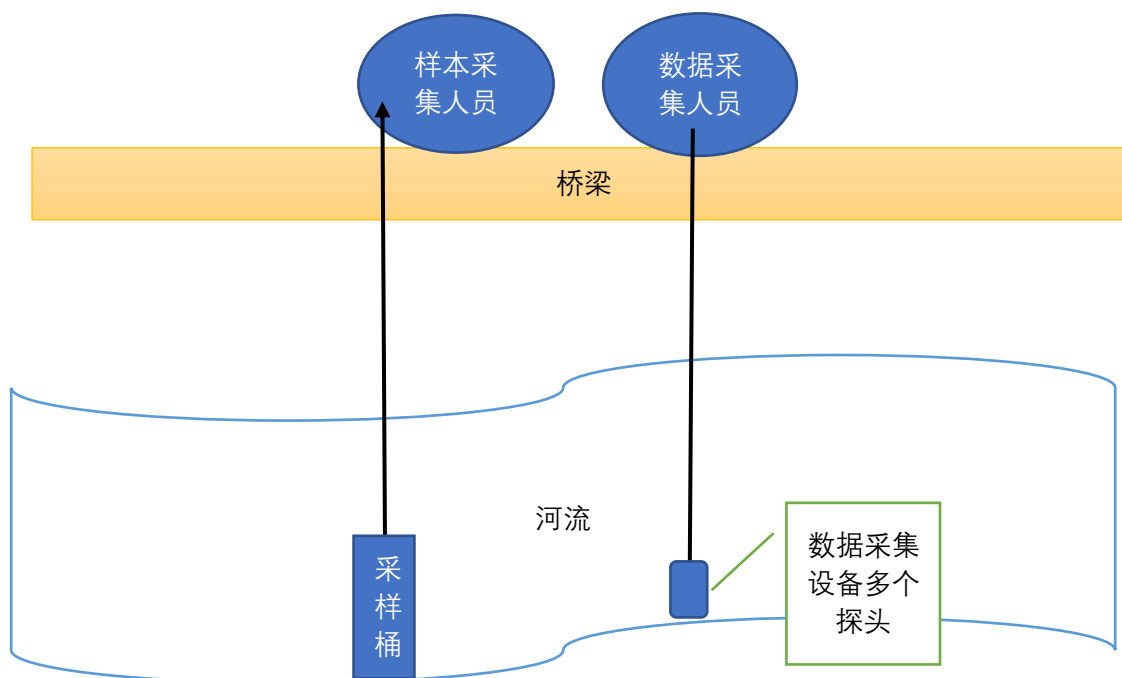
采样桶底部有可以活动的挡板，拉住绳子将采样桶放置于水下，其深度由河流的深度所决定。随后将采样桶拉上，用采样瓶在塑料软管处收集样品。（在收集样品之前务必将采样桶与采样瓶用所采集的水清洗两到三次）

第二类设备为数据采集设备，包括溶解氧仪，水温计等等（右图设备包括多个探头，可将所需的数据显示到屏幕上）。



(2) 采集方法

此次采集数据的河流断面大多都在各个桥梁，所以数据的采集基本都在桥梁上进行。具体方法如下图所示。



数据采集人员读取数据时，待各个数据显示趋近于稳定后再记录数据。样本及数据采集完毕后，用保护套防止探头被损坏。

1.2 记录，整合数据

根据数据采集设备，将其中对于水质分析有用的数据进行整合，绘制出表格。并结合历年来的数据，绘制出数据更丰富的表格以及柱状图，折线图等等。各类数据如下多个表格所示：

2014 年至 2019 年各断面水温情况

水温℃	2014	2015	2016	2017	2018	2019
龙王山	23.7	24.2	24.7	24.7	22.3	23.1
孝丰大桥	28.8	27.9	29.3	33.7	28.1	30.7
乌巷大桥	27.7	27.3	29.2	31.9	28.4	27.9
安城大桥	29.8	29.6	30.3	34.8	29.6	31.5
梅溪大桥	29.4	30.4	32	32.5	31.9	31.3
港口大桥	30.6	31.2	32.2	33.8	32.3	31.8
白雀大桥	31.6	32.8	32.5	32.3	31.2	30.5

2014 年至 2019 年各河流断面 pH 值情况

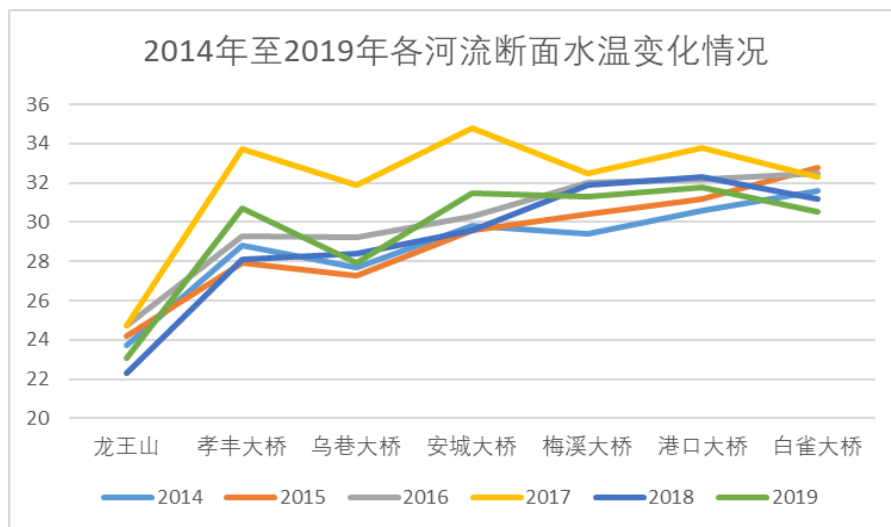
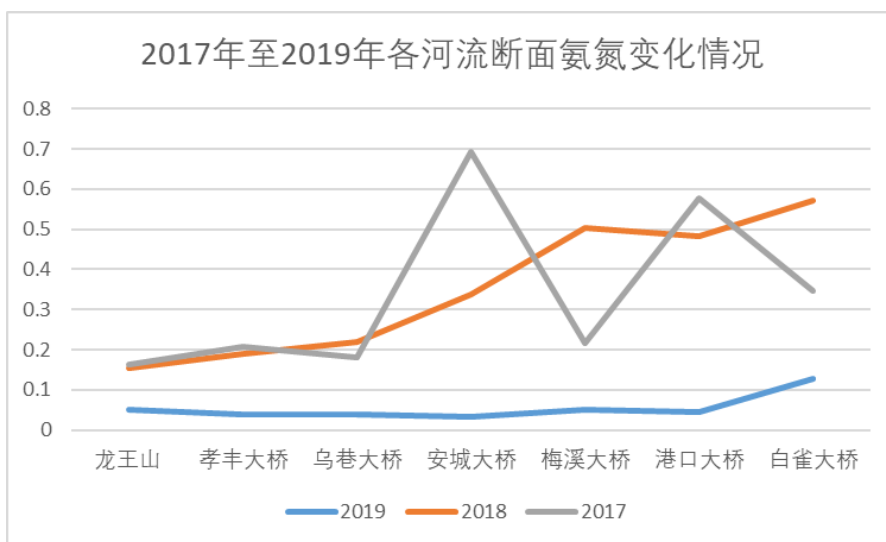
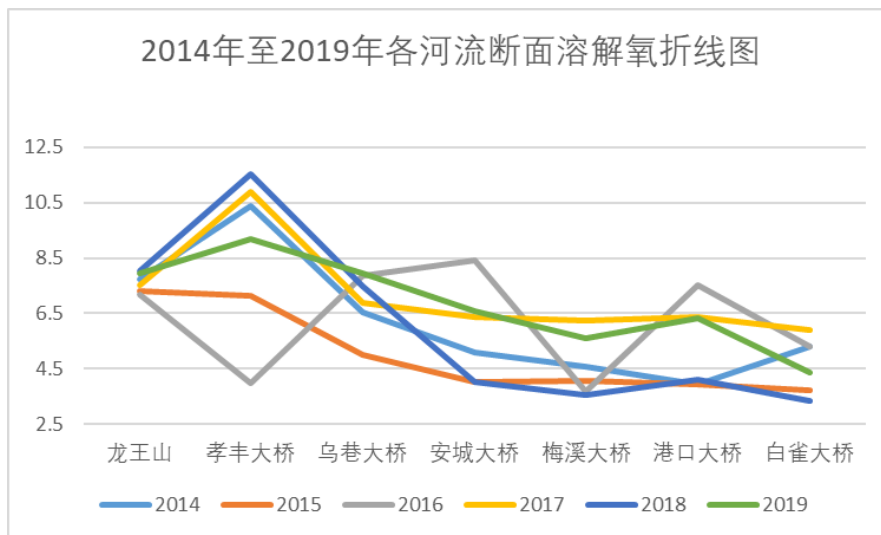
pH 值	2014	2015	2016	2017	2018	2019
龙王山	6.74	7.28	6.86	7.82	7.69	8.14
孝丰大桥	6.8	8.95	6.59	8.43	8.42	9.04
乌巷大桥	7.44	8.17	7.21	7.75	8.05	7.91
安城大桥	7.27	7.5	7.41	7.4	7.44	7.8
梅溪大桥	7.15	7.39	6.6	7.2	7.4	7.94
港口大桥	7.14	7.26	7.18	7.29	7.36	7.91
白雀大桥	7.56	7.72	7.26	6.87	7.35	8.52

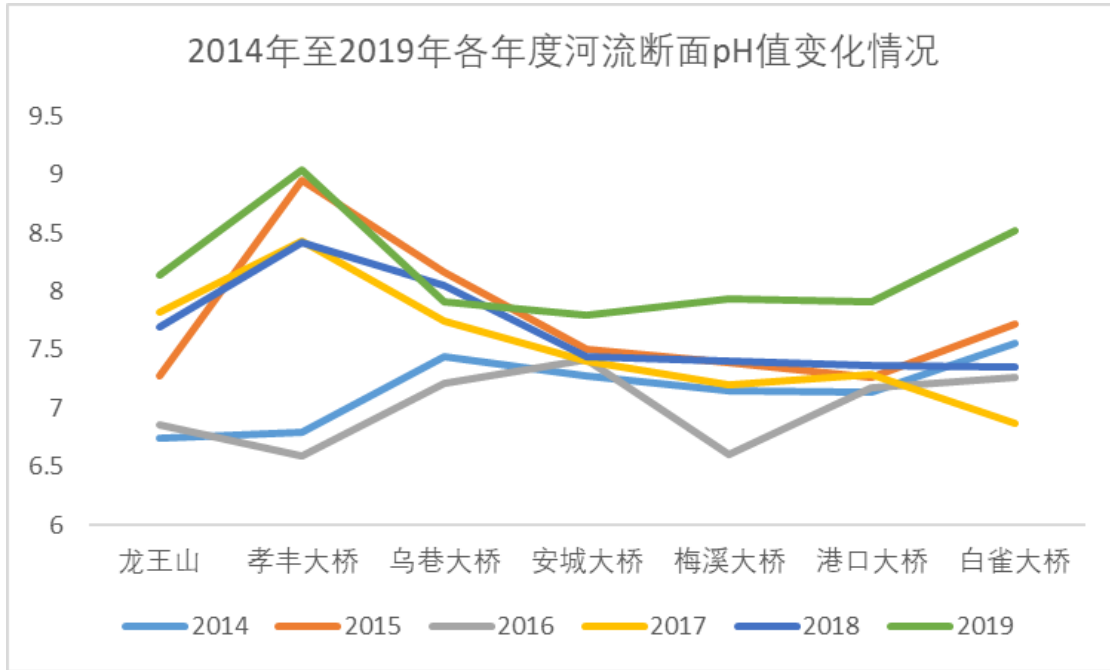
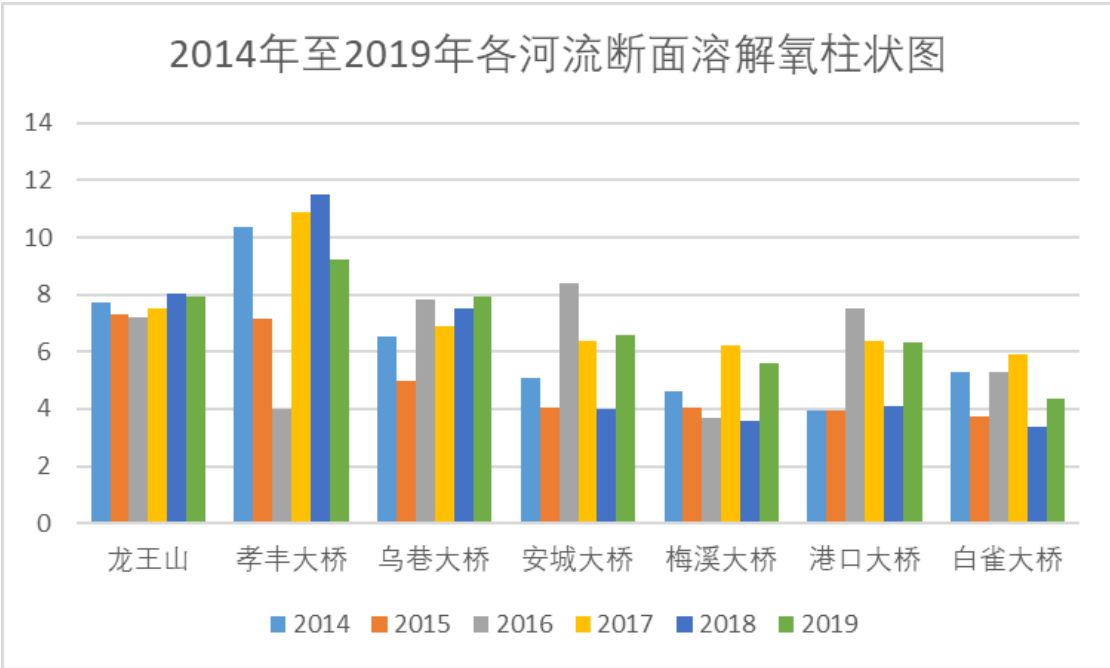
2014 年至 2019 年各河流断面溶解氧情况

溶解氧 mg/L	2014	2015	2016	2017	2018	2019
龙王山	7.72	7.3	7.2	7.52	8.05	7.95
孝丰大桥	10.36	7.15	4	10.9	11.52	9.2
乌巷大桥	6.55	5	7.85	6.89	7.5	7.95
安城大桥	5.08	4.05	8.41	6.38	4.01	6.57
梅溪大桥	4.6	4.07	3.7	6.23	3.56	5.59
港口大桥	3.93	3.94	7.51	6.38	4.1	6.32
白雀大桥	5.3	3.72	5.3	5.92	3.36	4.37

2017 年至 2019 年各河流断面氨氮、化学需氧量、五日化学需氧量情况（由于 2018 年化学需氧量信息不足，补充了 2016 年数据，对问题研究影响不大）

指标	时间	龙王山	孝丰大桥	乌巷大桥	安城大桥	梅溪大桥	港口大桥	白雀大桥
氨氮 (mg/L)	2019	0.052	0.038	0.038	0.034	0.05	0.044	0.127
	2018	0.155	0.191	0.218	0.337	0.504	0.482	0.573
	2017	0.163	0.208	0.182	0.694	0.216	0.577	0.348
化学需氧量 (mg/L)	2019	2.9	4.9	3.5	6.3	5.4	4.5	18.7
	2017	2.4	6.1	5.6	6.5	6.6	6.1	8.5
	2016	3.3	5.5	3.4	5.9	7.5	11.8	11.0
五日化学需 氧量 (mg/L)	2019	0.7	1.2	1	1.7	1.1	1.4	4.1
	2018	-	-	-	-	-	-	-
	2017	0.8	1.7	1.4	1.7	2	2	2.1





注：历年数据来源于黄浦江源头水质历年调查专项。

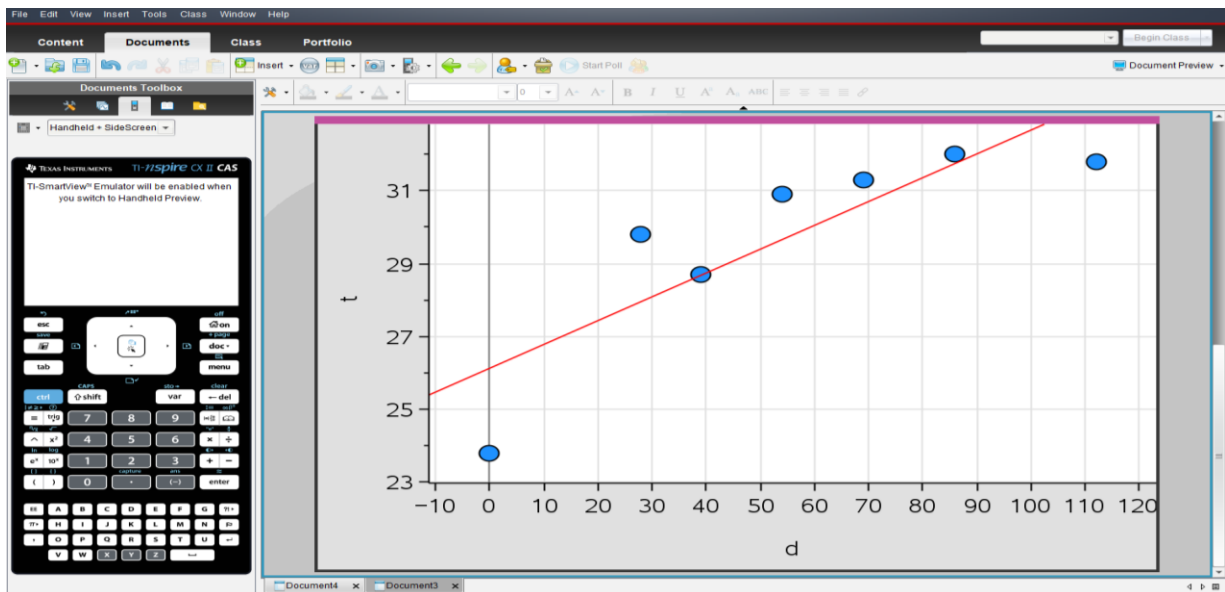
第二部分：数据的相关性分析及统计回归模型

本部分将分析上述数据各个因变量（response or independent variable）对自变量（predictor or independent variables）的相关性，以及分析如何建立合适的统计回归模型。为了研究方便，我们试图寻找各个指标和区域地理位置的关系。从本质上讲，这些相关关系可能是通过第三个变量（lurking variable）来互相关联的，例如可能是通过该区域的工业排放这个因素关联起来的。本文主要根据散点图（scatter plot）初步判断变量之间可能的相关关系，然后利用计算器和 Excel 计算相关统计量和回归模型，然后根据计算机输出报告（Summary out）中有关的假设检验（Hypothesis test）的结果对模型中的参数以及模型的拟合程度（goodness of fit）等进行分析和模型改进。

2.1 各区域水温变化回归模型

下面是对第一部分的数据做简单的计算，分别是各个采样点的水温以及他们相对龙王山的距离，根据散点图，我们看到越是远离龙王山的下游，水温越是相对高一些。所以我们试图用一次回归对样本做回归，把两组数据距离和平均水温输入计算器，其结果如下：

水温℃	距离的平方	到龙王山距离	平均
龙王山	0	0	23.8
孝丰大桥	784	28	29.8
乌巷大桥	1521	39	28.7
安城大桥	2916	54	30.9
梅溪大桥	4761	69	31.3
港口大桥	7396	86	32
白雀大桥	12544	112	31.8



	A d	B t	C	D	E
=					=LinRegB
1		0	23.8	Title	Linear R...
2		28	29.8	RegEqn	a+b*x
3		39	28.7	a	26.1358
4		54	30.9	b	0.065334
5		69	31.3	r ²	0.725744
6		86	32	r	0.851906
7		112	31.8	Resid	{-2.3357...

E1 = "Linear Regression (a+bx)"

根据上面的结果，回归系数 $r = 0.851906$ ，说明上述两组（样本）数据具有较强的相关性，水温 = $26.1356 + 0.065 \times$ 离开山顶龙王山距离。如果对上述线性模型做假设检验 ($H_0: \rho = 0, H_1: \rho \neq 0$)，即在原数据组没有相关关系的假设下获得各个统计量的可能性。

输出报告如下：

SUMMARY OUTPUT	
Multiple R	0.851906138
R Square	0.725744067
Adjusted R Square	0.670892881
标准误差	1.647718731
观测值	7

方差分析 (ANOVA)

	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
回归分析	1	35.92226	35.92226	13.23115	0.014943135
残差	5	13.57489	2.714977		
总计	6	49.49714			

根据上面的 T 和 F 分布检验，F 分布的 P-VALUE =0.0149, 其意味着只要显著性水平（Significance level）大于等于 1.5%，我们都可以拒绝原假设（两组数据没有相关性），换句话说，我们有足够证据表明二者具有线性相关性。

对于一元线性回归，通过 T 检验对方程系数的检验和对总的相关性（或相关系数）的 F 检验是等价的，即总体的相关系数 $\rho=0$ 是否为零和一次项系数是否为零的检验是等价的，所以就可以看到 Significance F =P-VALUE (T 检验)=0.014943。

T 检验：

	<i>Coefficients</i>	<i>标准误差</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	26.13579698	1.174313	22.25624	3.4E-06	23.11712926	29.1544647
距离	0.06533356	0.017961	3.637464	0.014943	0.019162587	0.11150453

对多个系数的 T 检验，将在对模型做稍微调整后再连同 EXCEL 输出报告中的信息一起做较为详细的说明。

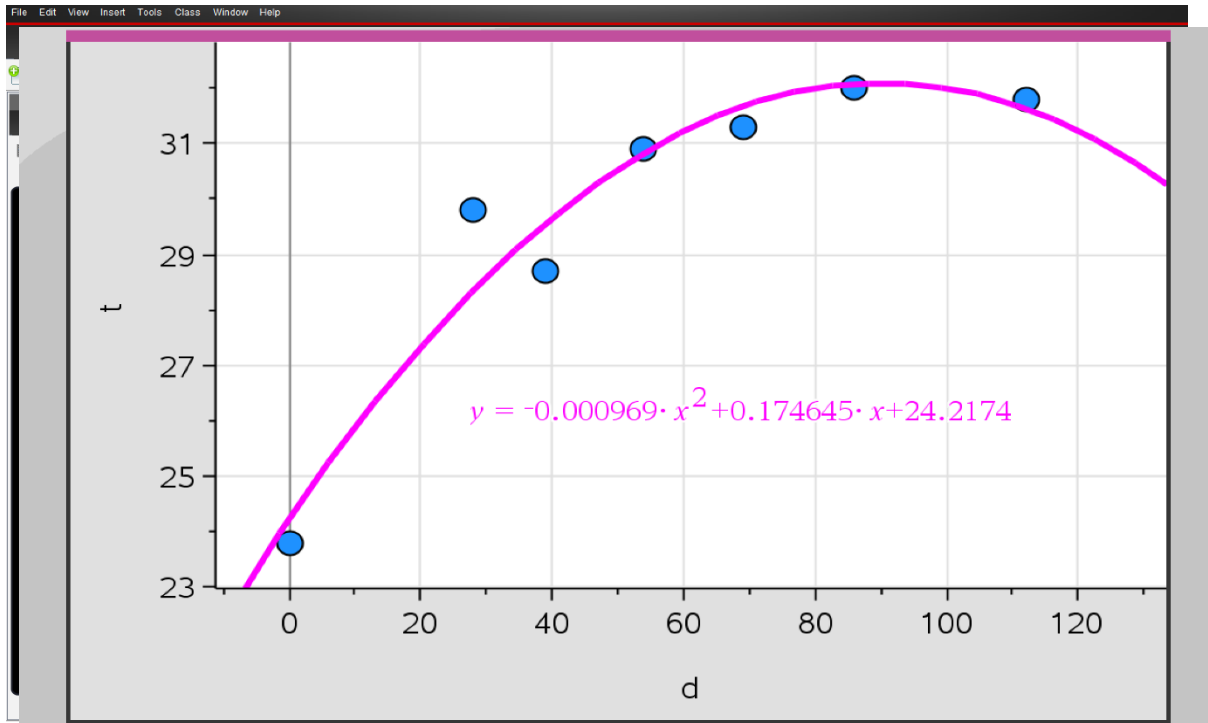
2.2 改进模型

虽然一次回归取得了可以接受的效果，但如果仔细观察散点图，我们就会发现回归直线可能还有改进的地方。如果利用二次多项式进行回归，曲线的拟合程度可能会更好。当然这还要看二次项的系数是否太小，以及检验报告结果是否认为有必要加入二次项系数。所以，我们将利用如下回归函数作为模型：

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad \varepsilon \text{ 代表误差，一般认为是服从 } N(0, \sigma^2) \text{ 的正态分布。}$$

这里用圆体的希腊字母代表总体的参数， $\beta_0, \beta_1, \beta_2$ 是所谓的“true”真实的系数。我们是用（可能每次不同的）样本来估计母体的参数。

这里我们做一点点技术处理就可以利用 EXCEL 里的二元线性来处理了，就是把距离的平方作为 X_2 来处理，然后在利用 EXCEL 运行回归的时候把距离的平方作为第二个自变量变量观察值。其结果和利用计算器直接选择 Quadratic 二次曲线回归结果完全一致。下面是回归结果及其结果分析：



Excel 回归模型报告：

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0.967325075
R Square	0.935717801
Adjusted R Square	0.903576701
Standard Error	0.891877962
Observations	7

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	46.31535766	23.15767883	29.112812	0.004132201
Residual	4	3.181785193	0.795446298		
Total	6	49.49714286			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	24.21744	0.82806	29.24590	0.00001	21.91837	26.51651
距离的平方	-0.00097	0.00027	-3.61466	0.02246	-0.00171	-0.00022
距离	0.17464	0.03177	5.49795	0.00534	0.08645	0.26284

2.3 回归分析报告的统计学说明，模型的假设检验

1. 上表是 EXCEL 回归分析的典型输出。整个报告分为三个部分：第一部分主要是有关统计量 R^2 ，即相关系数及其平方，反映的是两组样本的相关性大小。 R^2 square，也叫 coefficient of determination，是相关系数的平方，也等于 SSR/SST，是变量 Y 的变化多大程度上是由变量 X 的变化来解释的，所以 SSR 也叫 explained sum of square。SST 是总的 sum of square， R^2 愈大，说明因变量的变化用自变量的变化可解释的比例越高，相关性越高。
2. 第二部分是一个方差检验，是利用 F 分布来对整体的模型进行检验（overall usefulness of the model）。如果 F 检验失败，那第三部分的有关各个系数检验就没有必要，整个模型就要重来。F 检验的 $H_0: \beta_1 = \beta_2 = 0$ ， $H_1: \beta_1, \beta_2$ 至少有一个不为零，或 $H_0: R^2 = 0$ $H_1: R^2 > 0$ 二者实质相同。统计量 F 的计算公式：
$$F = \frac{R^2/K}{\frac{1}{n-K-1}(1-R^2)}$$
，在上面的报告里由平均和 MS (regression) 与 MS (residual) 的商获得，即 $23.15767883/0.795446298=29.112812$ 。最后的 P-Value = Significance F=Cdf(29.112812, ∞ , 2, 4)= 0.0041322012221856（本概率是通过计算器获得，和上面的 EXCEL 输出结果一致）。有关自由度的个数，可以参考相关的资料。
3. 第三部分是对一次项和二次项系数分别进行检验。对一般的多元回归而言，它代表各个自变量是否对（predicted variable Y）因变量有贡献，或二者相关性有多大做出检验，是利用 T 分布来检验的。如果某个系数为零，则该系数对应的自变量和因变量无关，可以从模型中去掉。例如，如果要做学习成绩与学习时间和体重的回归分析，很可能会发现学习成绩和体重没有相关关系，对于体重系数为零的假设可能不能得到充分的证据拒绝，也就是要维持原假设。
4. 上面 2.2 改进模型部分，对前面得出的线性回归方程增加了一个二次项作为变量，重新运行了数据。显然二次项的系数将会是很小的，说明它对因变量水温的贡献不大，但是仍然有必要，因为其系数检验的 P-Value=0.02246，说明原假设因变量和二次项系数没有关系，即二次项系数为零的概率是非常低的（低于显著性水平 3%），所以我们有充分的理由认为原假设不成立，应

- 该拒绝，也就是说二次项系数对因变量是有显著影响的。由此而得出的回归方程修正为 **水温=-0.000969*距离平方+0.174645*距离+24.2174**。
5. 从散点图看上面的二次(线性)回归要比一次回归对数据的拟合程度更好些。
 6. R 是样本回归系数 (Pearson's product moment correlation coefficient), 反映的是两组观测值的线性相关程度。定义和计算可以查阅一般的教科书。
 7. R 的平方是相关系数的平方, 由 $R^2 = SSr/SSt$ 来定义和计算, 可以证明和上面的皮尔森相关系数的定义是等价的, 其含义是因变量的变化多大程度上是由自变量的变化引起的, 在上表中 $SSr = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 46.31535766$ $SSst = \sum_{i=1}^n (y_i - \bar{y})^2 = 49.49714286$ 。
 8. $Standard\ Error = \sqrt{\frac{1}{n-k-1} SSe}$ 其中 $SSe = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = SSst - SSr = 3.181785193$
 $MS = SSE/df = 3.181785193/4 = 0.795446298$ $Standard\ error = \sqrt{0.795446298} = 0.891878$ 。
 9. 两个自变量的系数的 T 检验, $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ 其 T 统计量的计算方法为: $coefficient/standard\ error = t\text{-stat}$ 。由于上表在计算过程中做了近似, 我们采用原始数据, 二次项系数的 T 检验统计量为: $-0.000968662/0.000267982 = -3.61466$, $p\text{-value} = 2 * t\text{CDF}(-\infty, -3.61466, 4) = 0.022463419974624$ 。如果按照 5% 的显著性水平, 应该拒绝原假设: $\beta_1 = 0$, 其逻辑和任何假设检验一样: 当我们在以原假设为真的前提下, 计算在一次试验中获得试验样本的概率, 包括比样本更极端的情况, 如果此概率非常小, 小于我们预先设定的标准 (显著性水平), 那么, 我们就认为我们有充分的证据拒绝原假设。这里, 原假设是系数为 0。
 10. 最后, 有关统计量 $F = \frac{R^2/K}{\frac{1}{n-K-1}(1-R^2)}$ 和 统计量 $T = \frac{R}{\sqrt{\frac{1-R^2}{n-2}}}$ 分别服从自由度为 $F(k, n-k-1)$ 和 $T(n-2)$ 。

2.4 对各个区域 PH 值和溶解氧的相关关系的探讨

根据上面的数据, 我们观察到 PH 值可能和溶解氧有一定的相关关系。它们都在孝丰大桥附近得到峰值, 然后到下游有所下降, 但 PH 值下降不明显, 下游基本上为中性, 而溶解氧下游变得非常小, 两者虽然有一定的相关性, 但不够明显。下面对数据作回归分析, 看看我们的直觉是否正确。

平均PH值	溶解氧
7.4217	7.6233
8.0383	8.8550
7.7550	6.9567
7.4700	5.7500
7.2800	4.6250
7.3567	5.3633
7.5467	4.6617

SUMMARY OUTPUT	
回归统计	
Multiple R	0.7520
R Square	0.5655
Adjusted R Square	0.4786
标准误差	0.1896
观测值	7

方差分析

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
回归分析	1	0.2339	0.2339	6.5065	0.0512
残差	5	0.1797	0.0359		
总计	6	0.4136			

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	6.7802	0.3112	21.7873	3.7804E-06	5.9802	7.5801
溶解氧	0.1234	0.0484	2.5508	0.0512	-0.0010	0.2477

根据上述输出报告，对相关性 R^2 的检验处于尴尬水平。F 检验 $P\text{-VALUE}=T$ 检验 $=0.051216817$ ，结果和我们的直觉一致，即虽然二者有一定的相关性（也许是通过区域中其他第三变量）但相关关系不明显。从原始数据看，溶解氧变化较大，不能通过 PH 值的微小变化来解释，需要将来更多的观察和数据来探究其变化规律。

2.5 平均溶解氧回归模型：

根据 2.4 部分的数据，平均溶解氧在上游较高，随着河流向下很快降低，具有显著的变化规律。我们根据散点图，很容易看到其变化规律。下面对数据运行回归分析：

从散点图看，曲线并不像水温变化那么有规律，这和 excel 输出报告的 R 值一致，但按照回归分析报告，P-value 都很小，说明相关关系非常显著。Adjusted $R^2=0.62$, 表明溶解氧的变化的 62% 可以用地理位置的变化来解释。从生态环境的变化角度来看，可能随着河流向下，有新的水源汇入，稀释了溶解氧的浓度。按照类似的分析，我们可以建立溶解氧随区域位置变化的规律：**溶解氧=8.22145-0.035348*离开龙王山距离**。同时可以看到, 计算器和 EXCEL 输出的回归方程也完全一致。据此模型我们可以对区域内任意位置的溶解氧做估算，然后和实际数据做比较来监测环境是否有显著的变化。

第三部分：从环境科学对统计模型的定性验证

本文上一部分是在分析，研究第一部分原始数据的基础上对与环境有关的几个变量做了定量分析，建立了数学模型，并对模型进行了统计学假设检验。本部分将从环境科学的角度对水温，溶解氧，PH 值的变化做定性的因果分析，把统计模型再一次和现实环境做比较，其变化趋势和定量分析一致。

在本次黄浦江上游水质考察项目中，本人发现安吉县有大量的竹制品生产工厂。返回后使用网络地图，在各个河流断面周围进行搜寻，发现了孝丰东门大桥、乌巷大桥两河流断面周围都有各类竹制品生产工厂，如下图所示。并且不仅如此，在龙王山到孝丰大桥河断面处，河流两岸都出现了大量的竹制品厂。其排出的工业废水对下游的水质肯定有着不小的影响。



3.1 分析并对比孝丰大桥和乌巷大桥河流断面数据

(1) 水温

通过分析同年度的各个河流断面的水温，见 2014 至 2019 年各个断面水温变化情况折线图，发现两河流断面及离它们较近的安城大桥的水温较其它河流断面水温偏高，其中孝丰大桥更为显著。由于该两个河流断面的地理位置与其它五个河流的位置相差不远，各个河流断面周围的空气温度应相差不大（除了龙王山山顶），而水温却有着大约三度的温差。

以下为造成此温差的原因的猜想：

由公式：

$$c = \frac{Q}{m \cdot \Delta t}$$

其中，c 为比热容（比热），是物质的一种特性，Q 表示单位时间内物体吸收或放出的热量，m 为物体质量， Δt 为时间。

经过实验测定，水中若溶有钾盐，钠盐等其它盐类物质，会导致其水溶液的比热容减小，且浓度越大比热容越小。可能是由于在乌巷大桥以及孝丰大桥河流断面处或龙王山到孝丰大桥的河段有含有盐类的废水排出，导致其下游水中盐类浓度提升，比热容减小，温度相对较高。

(2) pH 值

通过分析对比同年度的各个河流断面的 pH 值，观察折线图，可以发现多数年度的河流断面 pH 值甚至是在孝丰大桥达到了峰值（碱性），并在向下游河流断面处流动时缓缓稀释，pH 值逐渐靠近 7（中性）。

原因猜想：

由于孝丰大桥以及乌巷大桥为龙王山水源地的后两个河流断面，在龙王山到孝丰大桥之间，或者是孝丰大桥河流断面周围，有偏碱性的废水排出，导致了孝丰大桥河流断面的 pH 值急剧升高，达到峰值。

(3) 溶解氧

通过分析对比同年度的各个河流断面的溶解氧，观察折线图及柱状图，可以发现大多年度的各个河流断面的溶解氧也是在孝丰大桥达到了峰值，最后逐渐减小。

溶解氧（DO）是水体经过与大气中的氧气交换或经过化学、生物化学等反应后溶解于水中的氧，一定的水中溶解氧的含量与空气中氧的分压、水温、水的深度、水中各种盐类和藻类的含量以及光照强度等多种条件有关。同时溶解氧也是反应了

细胞新陈代谢等重要生命活动基本参数之一。

原因猜想：

可能由于氧气在龙王山山顶到孝丰大桥河流断面之间的河段中不断地溶于水，便在孝丰大桥及乌巷大桥处达到峰值，在往下游的过程中水中的氧气不断被消耗。或可能是由于水中的藻类植物繁殖，进行了光合作用，导致水中的溶解氧增多，而在下游呈下降趋势。

从上面对所考察区域环境状况的定性分析可知，各项指标变化趋势和我们定量分析得出的结果相符合。

第四部分：结论

本文通过对实地考察，收集数据，统计分析，建立统计模型模型等一系列的过程，得出如下初步结论：

1. 考察区域水温变化模型： $\text{水温} = -0.000969 \times \text{距离平方} + 0.174645 \times \text{距离} + 24.2174$
2. 考察区域溶解氧变化模型： $\text{溶解氧} = 8.22145 - 0.035348 \times \text{离开龙王山距离}$
3. 考察区域 PH 值、溶解氧的相关性，及二者与周边生态环境及工业排放的相关性。
4. 探讨了如何将一元简单回归，一元高次多项式回归，多元线性回归及模型检验方法应用于环境科学的研究中，取得了预期效果。

致谢

感谢对本课题做出帮助的指导老师与同学们！

参考文献

- [1] 李党生,《环境保护概论 (第二版)》[M]. 中国环境出版社. 2013.
- [2] 常颖, 吴强主编.《净水工艺》[M]. 华南理工大学出版社. 2014.
- [3] 郑重.《安吉竹制品生产废水处理工艺研究--以逢春污水处理厂为例》[D]. 浙江: 浙江工业大学. 2015.