

## 研究区域患病情况介绍

### 1. 资料来源

基于云南省 2010-2015 年出生缺陷个案数据及出生数据 (共计 129 个县 (区)), 选择常住云南省的孕产妇所分娩的胎儿, 婴儿, 要求婴儿的出生日期 2010 年 1 月 1 日---2015 年 12 月 31 日

出生缺陷婴儿对象: 妊娠满 28 周至出生后 42 天内首次确诊的出生缺陷儿 (包括活产、死胎和死产)

### 2. 数据处理与空间化

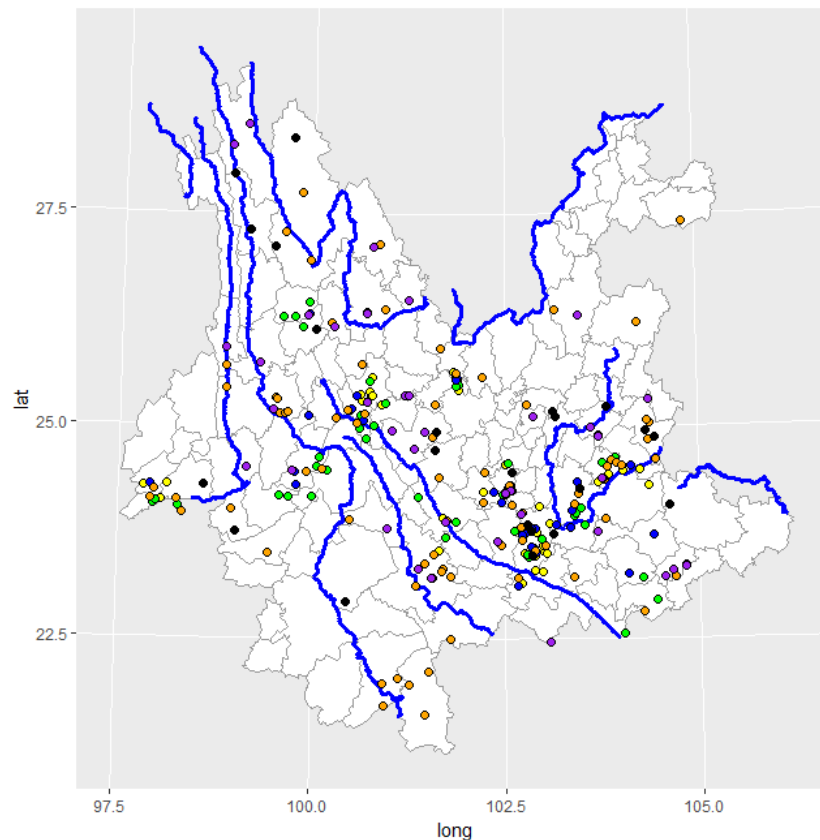
将云南省 2010-2015 年的出生数据和出生缺陷个案数据按照街、乡镇名进行分类汇总, 统计出各个县 (区) 在 2010-2015 年的出生缺陷例数和新生儿数

### 3. 出生缺陷计算

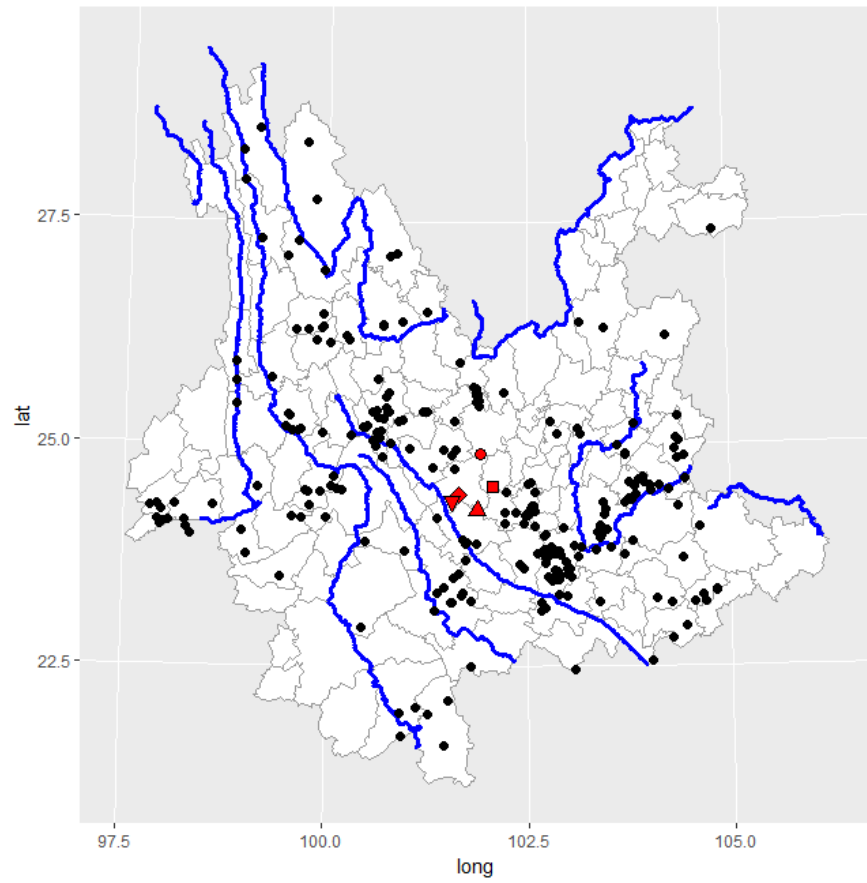
出生缺陷率 (1/万) = (出生缺陷数/出生总人数) × 10000

## 疾病点模式分析

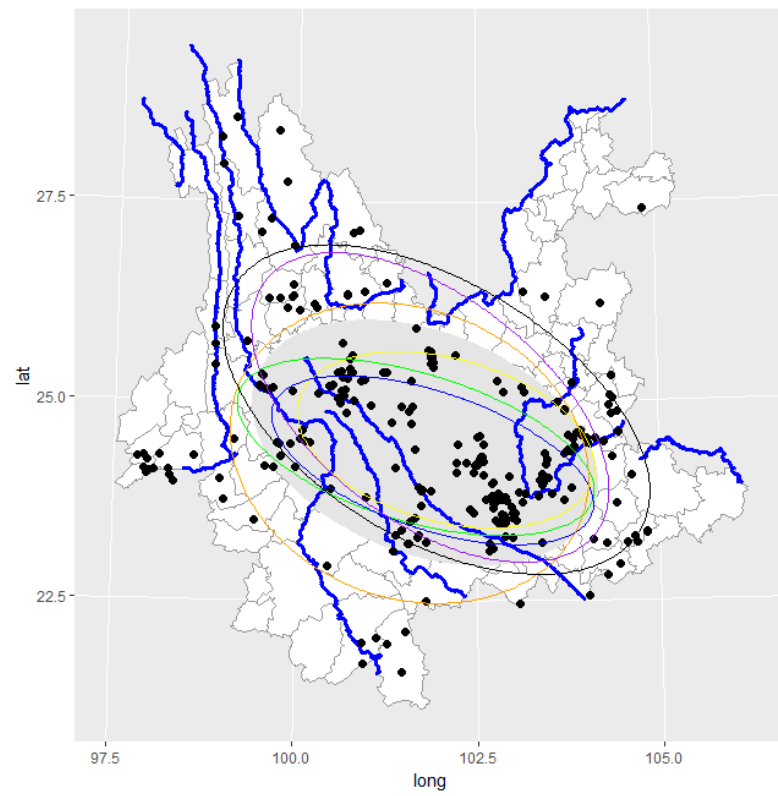
### 1. 历年的出生缺陷疾病分布



### 2. 历年的出生缺陷分布中心



### 3. 点分布的空间离散模式

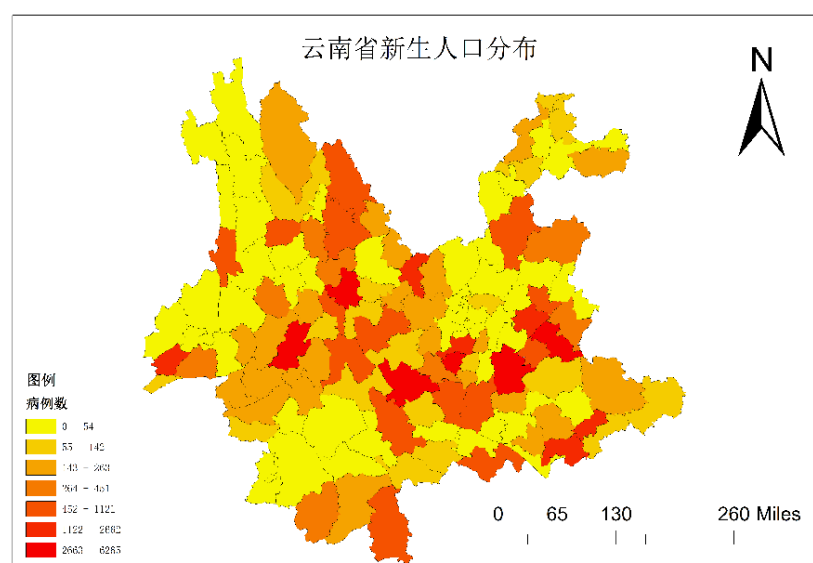
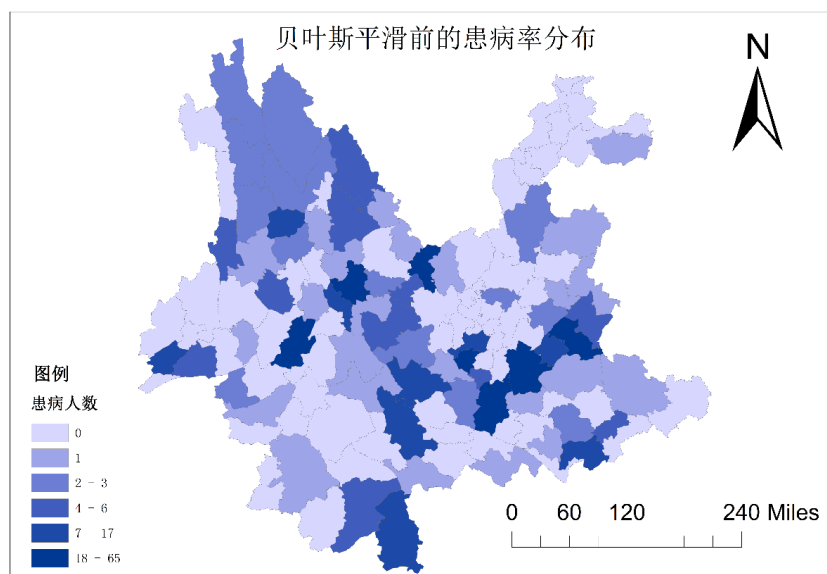


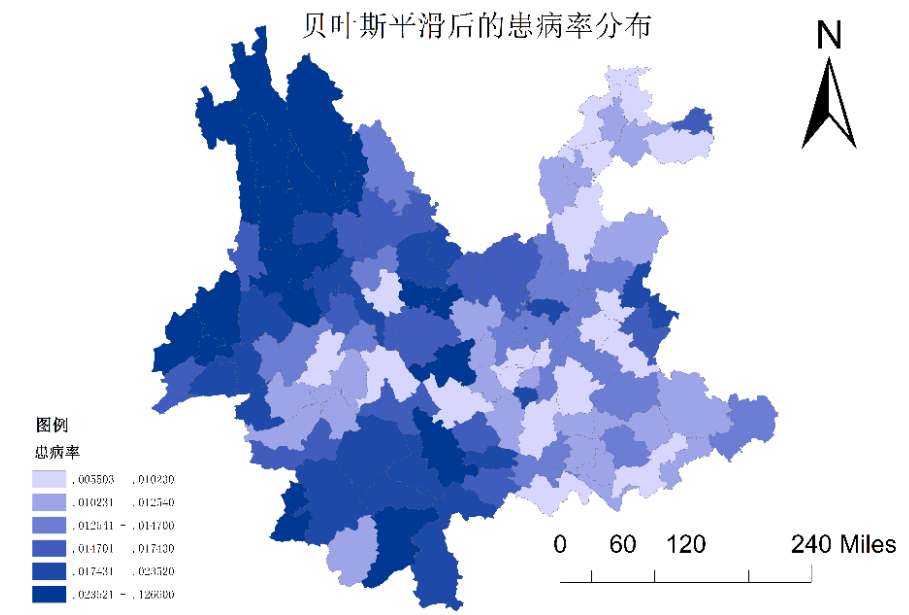
### 贝叶斯平滑 (binomial-logistic modeling)

各地区的患病人数服从二项分布，  
考虑到不同地区之间在空间上的相互影响，利用 logistic 模型获得平滑后的患病率  
具体公式：

$$Y_i \sim \text{binomial}(E_i \theta_i)$$
$$\text{logit } \theta_i = \beta + V_i + \varepsilon_i$$
$$\text{logit } \theta_i = \log\left(\frac{P_i}{1 - P_i}\right)$$

$V_i$ 表示了空间结构的影响，在 Winbugs 中由 Car.normal 公式实现，空间权重矩阵利用 Winbugs 中的 Adjacency Tools，对于相邻地区权重值取 1 否则取 0.， $\beta$ 代表出生缺陷的患病率值具有 uniform prior， $\varepsilon_i$ 表示了随机噪声数据具有 gamma prior.





### 莫兰指数计算

构建空间权重矩阵：基于相邻关系的空间权重矩阵

结果分析：

Moran I statistic standard deviate = 9.4134, p-value < 2.2 e-16

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.411335955	-0.007812500	0.001982643

### 地理加权回归模型

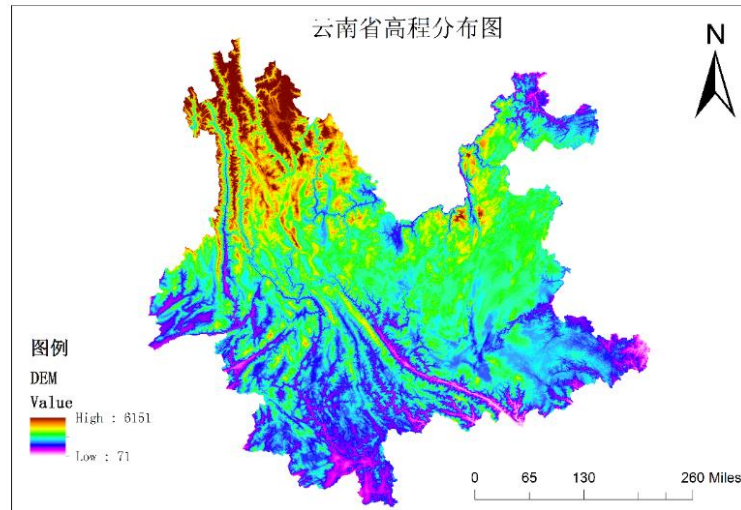
#### DEM

数据源： SRTMDEM 90m 高程数据， 患病数据

数据预处理： 计算出每个县级地区的平均高程与患病数

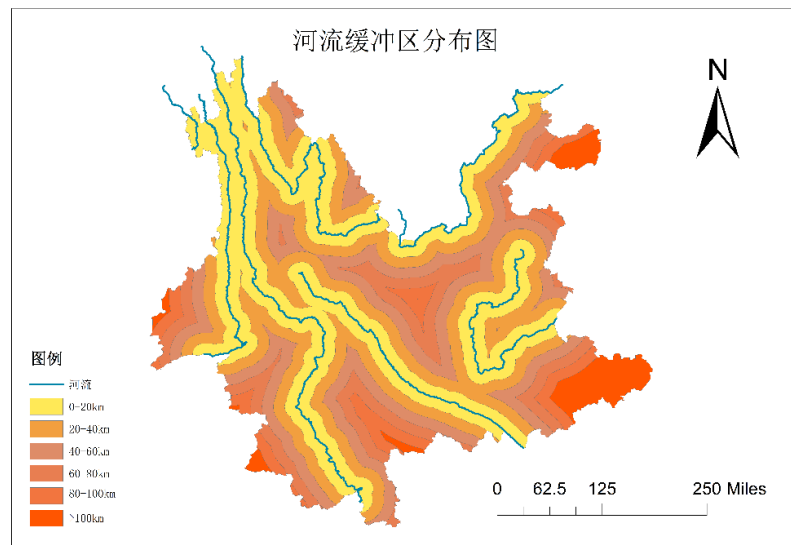
结果： 区域的高程值与当地的出生缺陷患病率之间存在一定的正相关性即在高程较高的地区， 新生儿患出生缺陷的可能性较大

如下图为云南地区的高程分布图



### 河流缓冲区

现根据云南地区的六大河流建立不同范围的缓冲区，缓冲区如下图所示，分析出生缺陷患病率与地区距离河流的位置之间是否存在相关性。缓冲区分为（1: 0-20km, 2: 20-40km, 3: 40-60km, 4: 60-80km, 5: 80-100km, 6:>100km）



### 地区内通婚率

数据：利用新生儿父母双方的身份证前六位的区域编号来对父母家乡与区域进行匹配，从而计算出各个县（区）内部的通婚率

### 地区的 GDP 值

数据：云南省统计局统计年鉴

a) 利用蒙特卡洛模拟分析获得各个地理因素的 P-value 如下

Independent value	DEM	buffer	marriage	GDP	intercept
P-value	0.00	0.05	0.00	0.05	0.5

b) 地理加权回归分析结果

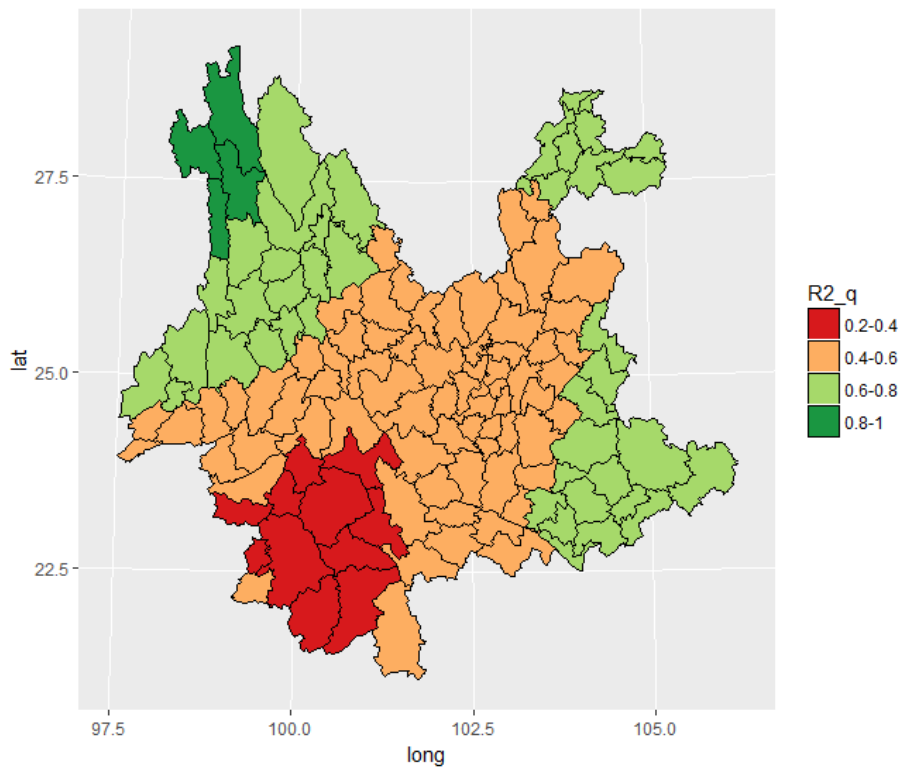
```

*****
*               Results of Geographically Weighted Regression               *
*****

*****Model calibration information*****
Kernel function: gaussian
Fixed bandwidth: 1.509172
Regression points: the same locations as observations are used.
Distance metric: A distance matrix is specified for this model calibration

*****Summary of GWR coefficient estimates:*****
              Min.      1st Qu.      Median      3rd Qu.      Max.
Intercept -1.6273e-02  1.1836e-03  1.0753e-02  1.5462e-02  0.0380
DEM        -6.5532e-06  1.8977e-06  3.4230e-06  1.1042e-05  0.0001
buffer     -6.6124e-03 -2.1040e-04  6.8291e-05  4.8648e-04  0.0016
marrige    -1.2917e-01 -6.1791e-03 -3.6322e-03 -6.9391e-04  0.0053
GDP        -7.4916e-07 -9.6371e-08 -5.2736e-08 -3.9041e-08  0.0000
*****Diagnostic information*****
Number of data points: 129
Effective number of parameters (2trace(s) - trace(s's)): 22.95286
Effective degrees of freedom (n-2trace(s) + trace(s's)): 106.0471
AICC (GWR book, Fotheringham, et al. 2002, p. 61, eq 2.33): -887.5342
AIC (GWR book, Fotheringham, et al. 2002, GWR p. 96, eq. 4.22): -913.5124
Residual sum of squares: 0.005545743
R-square value: 0.7408692
Adjusted R-square value: 0.684249
*****

```



### 非环境因子分析

相关因素	调查人数	患病数	患病率 (%)	$\chi^2$	P 值
母亲文化程度					

文盲	968	10	1.03	13.447	0.01953
小学	16599	130	0.783		
初中	36199	211	0.563		
高中	7676	42	0.547		
大专/大本	3715	16	0.431		
研究生及以上	29	0	0		
父亲文化程度					
文盲	337	2	0.593	10.88	0.05382
小学	15539	125	0.804		
初中	38264	217	0.567		
高中	7115	44	0.618		
大专/大本	3656	19	0.520		
研究生及以上	21	0	0		
母亲职业					
农民	63560	400	0.629	3.3061	0.7696
工人	250	0	0		
服务业	370	2	0.541		
经商	197	1	0.508		
家务	113	1	0.885		
教师/公务员	529	3	0.567		
其他	350	4	1.143		
父亲职业					
农民	62641	395	0.631	3.8335	0.6992
工人	400	3	0.75		
服务业	329	0	0		
经商	297	2	0.673		
家务	12	0	0		
教师/公务员	508	1	0.197		
其他	540	3	0.556		
母亲民族					
汉族（1）	40090	232	0.579	3.4976	0.06149
少数民族（0）	25725	180	0.7		
父亲民族					
汉族	42665	236	0.553	10.416	0.001249
少数民族	22580	173	0.757		
母亲患慢性疾病				1.9453	0.1631
无	65069	404	0.621		
有	972	10	1.029		

母亲患地中海贫血症					
是	65603	412	0.628	0.02281	0.8813
否	438	2	0.457		
父亲年龄					
小于 35 岁	60196	358	0.595	9.3964	0.002174
大于 35 岁	4997	48	0.961		
母亲年龄					
小于 35 岁	63359	387	0.611	5.7683	0.01632
大于 35 岁	2445	25	1.022		

母亲孕期情况：

相关因素	调查人数	先天性缺陷数	畸形率 (%)	$\chi^2$	P 值
进食肉蛋类					
否 (0)	1681	8	0.476	0.40965	0.5235
是 (1)	64360	406	0.631		
厌食蔬菜					
否 (0)	64465	408	0.633	1.1919	0.207
是 (1)	1576	6	0.381		
妻子吸烟					
不吸烟 (0)	64602	398	0.616	13.019	0.004595
是 (1)	284	0	0		
减少 (2)	110	1	0.909		
不变 (3)	1045	15	1.435		
增加 (4)	0	0	0		
饮酒					
不饮酒 (0)	64528	396	0.614	15.836	0.001225
戒酒 (1)	391	1	0.256		
减少 (2)	95	2	2.105		
不变 (3)	1027	15	1.461		
增加 (4)	0	0	0		
孕妇患病毒性肝炎					
否(0)	66015	413	0.626	0.7015	0.4023
有 (1)	26	1	3.846		
丈夫吸烟					
0 (不吸烟)	29018	169	0.582	12.708	0.0128
1 (是)	5051	21	0.416		
2 (减少)	18974	117	0.617		
3 (不变)	12989	107	0.824		



4 (增加)	9	0	0		
接触农药	2014 年数据缺少				
无 (0)	59996	381	0.635	3.1479	0.07602
有 (1)	313	5	1.597		
接触家畜	2014 年数据缺少				
无 (0)	59526	378	0.635	1.26	0.2616
有 (1)	783	8	1.022		