

实验13、医学和流行病学统计

年级：15级 专业：生信 学号：1513401013 姓名：郑磊

编号 一 二 三 四 总分 评阅人

得分

软硬件平台：

1. 硬件平台：（硬件配置）i5，2.9HZ处理器，16G内存，64位操作系统
2. 系统平台：（操作系统及其版本号）Windows10 企业版
3. 软件平台：（软件系统及其版本号，若是在线分析平台，还需要提供URL地址）R3.4.1 ， Rstudio

一、目的要求：

- 1、加深对医学统计中常用统计量的理解；
- 2、熟悉并掌握医学统计中常用统计量的计算方法。

二、实验内容：

- 1、根据《2013中国卫生统计年鉴》内容，提取并可视化以下类型统计数据：

- (1) 不同年份和/或不同地区的医疗资源比较；
- (2) 不同年份和/或不同地区的人口出生率、死亡率比较；
- (3) 不同年份和/或不同地区的总体患病率比较；
- (4) 不同年份和/或不同地区的某种疾病患病率比较；每一个类别至少提取一组数据进行可视化对比（数据自行选择、可视化方案自行拟定），并对这些结果进行分析讨论。

- 2、中国国内心血管疾病、肿瘤等复杂疾病发病风险的调研资料阅读和分析：

2.1、文献检索：

选择一个感兴趣的疾病类型，在万方数据、中文期刊网（CNKI）、维普数据等文献数据库中，搜索该疾病以及遗传多态性相关的文献，从中任选一篇；阅读摘要或全文并提取出构建R*C列联表。

文献名：肿瘤坏死因子-β基因G252A、C804A多态性与冠状动脉粥样硬化性心脏病的相关性

2.2、对其中的等位基因频率和/或基因型频率进行卡方检验，并选取其中一组有显著差异的列联表数据进行后续分析。

C804A genotype frequency (C804A 基因型频率)			Allele frequency (等位基因频率)	
CC	CA	AA	C	A
0.257(54)	0.495(104)	0.248(52)	0.505(212)	0.495(208)
0.371(69)	0.457(85)	0.172(32)	0.599(223)	0.401(149)
	7.073		7.146	
	0.029		0.008	

2.3、如果原始列联表数据不是2*2类型，则根据数据分布特征转成2*2类型；假定疾病风险因素为暴露因素（E+），而后计算风险比（RR）、几率比（OR）、使用概率计算的几率比（OR'）。

2.4、样本量的计算：

调查CHD组中CC基因型的比例（± 5%区间）， α=0.05， π=0.9

$$n = (1.96/0.05)^2 \times (0.9 \times 0.1) = 138.2976 \approx 139$$

调查高血压组中IV基因型的比例（± 2.5%区间）， α=0.05， π=0.9

$$n = (1.96/0.025)^2 \times (0.9 \times 0.1) = 553.1904 \approx 554$$

=》要求比例越接近真实情况，样本量需求越大

调查对照组中IV基因型的比例（± 2.5%区间）， α=0.05， π=0.5

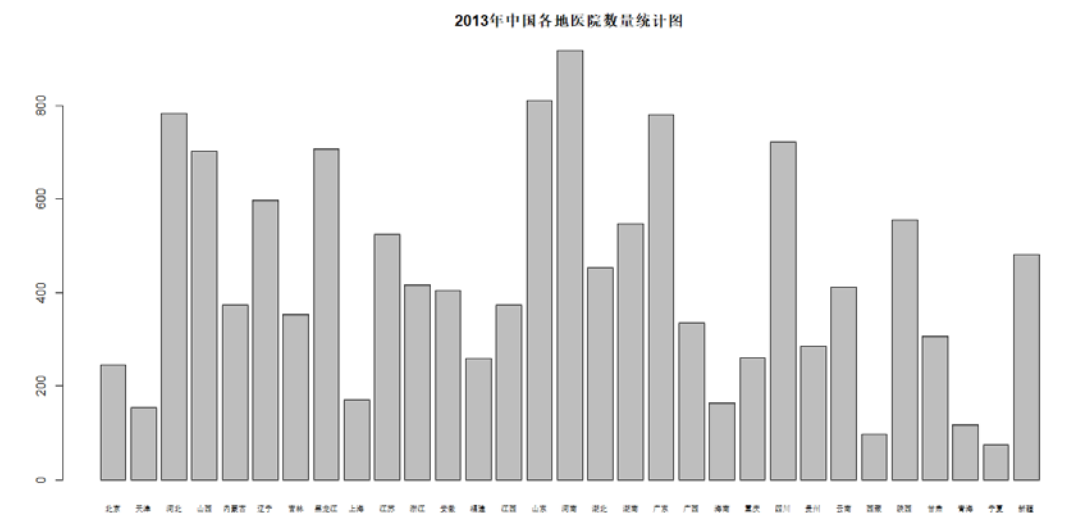
$$n = (1.96/0.025)^2 \times (0.5 \times 0.5) = 1536.64 \approx 1537$$

=》不同基因型比例分布越均匀，样本量需求越大

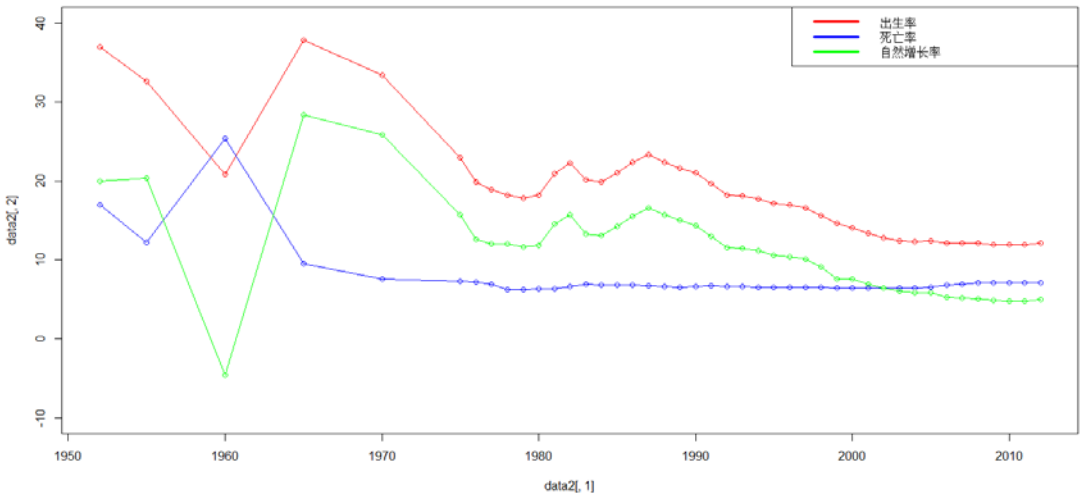
三、实验结果：

1、

(1)

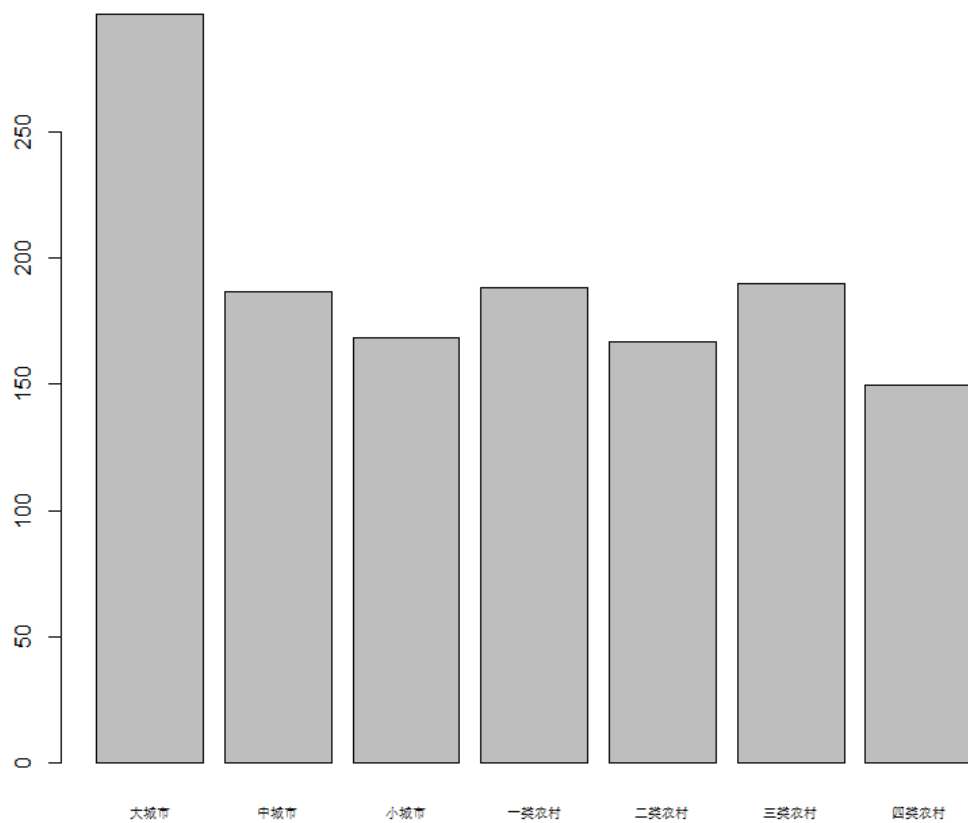


(2)



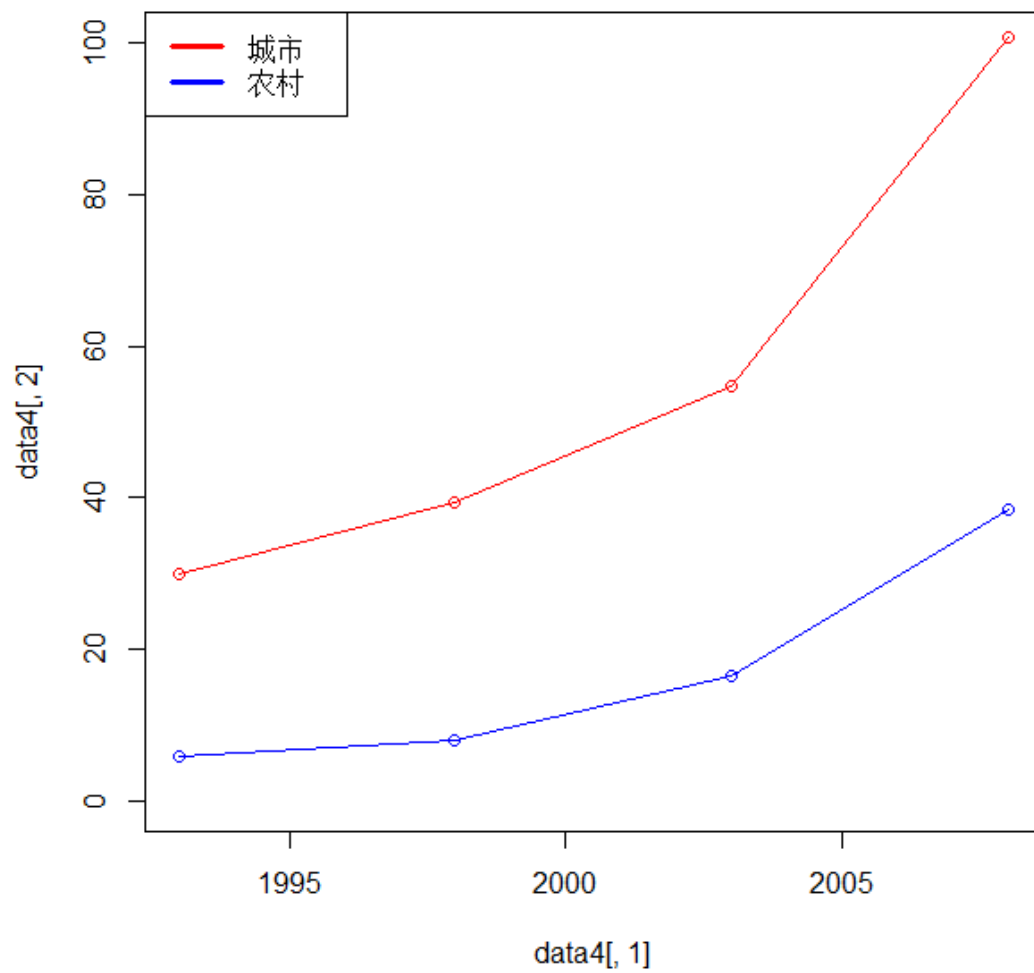
(3)

2008年中国患病率统计图



(4)

城市农村高血压患病情况



2.

2.2

```
> x<-cbind(c(212, 208), c(223, 149))
> chisq.test(x, correct=F)
```

Pearson's Chi-squared test

data: x

X-squared = 7.146, df = 1, p-value = 0.007513

```
>
```

```
> x<-cbind(c(54, 104, 52), c(69, 85, 32))
```

```
> chisq.test(x, correct=F)
```

Pearson's Chi-squared test

data: x

X-squared = 7.0727, df = 2, p-value = 0.02912

2.3

	CHD组	对照组	合计
CC	54	156	210
CA+AA	69	117	186
合计	123	273	396

$RR = (54/210)/(69/186) = 0.69317$

$OR = (54 \times 117)/(156 \times 69) = 0.88796$

$p1 = p(D+|E+) = 54/210 = 0.257 = \gg odds1 = p1(1-p1) = 0.191$

$p2 = p(D+|E-) = 69/186 = 0.371 = \gg odds2 = p2(1-p2) = 0.233$

$OR' = odds1/odds2 = 0.820$

四、讨论：

1

(1)

可以看出医疗资源的分配和城市人口有关，也和地区的经济水平有关，但总体来说，分配还是比较均衡的。

(2)

从1952年到2012年我国的人口出生死亡率，除了1960年附近出现了不正常的波动外，都很平稳，而由于计划生育国策的事实，我国人口出生率一直在下降，

人口净增长也一直在下降，虽然人口总数，但不可避免的迎来了人口老龄化。

(3)

可以看出，除了大城市，其余地方的患病率都十分接近，可能和大城市的生活节奏快，空气质量等有关。

(4)

可以看出，城市的高血压患者数量一直居高不下，最近几年增长速度尤其的快，已经达到了农村的好几倍，可能和缺乏运动和饮食习惯有关。