

基于 R 语言的生物统计学实战案例

实战内容
实战 1-统计数计算
实战 1-概率分布
实战 3-u 检验 t 检验
实战 4-样本频率的检验
实战 5-卡方检验和方差分析
实战 6-线性回归与相关分析
实战 7-作图-箱线图-柱状图-折线图-线性回归直线图

说明：

1. 本文档只含案例原题。多数案例取自：生物统计学，李春喜，科学出版社，2018 年（第五版）。

2. 基于 R 语言的实战代码、答案及数据见：

<https://github.com/gqliu1010/Biostatistics-R-implementation>

实战 1-特征数计算

例 1：产生 1-10 的整数，计算其平均值，中位数、众数、极差、方差、标准偏差、和变异系数。

例 2：求以下数列 1，2，5，7，9 平均值，中位数、极差、方差、标准偏差、和变异系数。

例 3：读入文本文件（file1_GC.txt），并计算每列平均值。

例 4：读入文本文件（file2_GC.txt），并计算每列平均值，并将结果存入文本文件（如 aver_GC.txt）

实战 2-概率分布函数

例 1:

(1)计算 10 次重复抛掷硬币试验中, 精确找到 3 次正面朝上的概率。

(2)在 50 次抛硬币试验中, 获得 0—25 次正面朝上的累积概率为多少?

(3)假设每次抛骰子点数 1 出现的概率为 $1/6$, 每组实验由 100 次抛骰子试验组成, 试在 0—100 范围内产生 10000 个服从二项分布的整数, 并作频数分布图。

例 3.4: 豌豆的红花纯和基因型和白花纯和基因型杂交后, 在 F_2 代红花植株与白花植株出现的比例为 3: 1。若每次随机观察 4 株, 共观察 100 次, 问得红花为 0 株、1 株、2 株、3 株和 4 株的概率各是多少?

例 3.5: 某批鸡种蛋的孵化率是 0.9, 从该批种蛋种每次任选 5 枚进行孵化, 试求孵出小鸡的各种可能的概率。

例 3.6: 某小麦品种在田间出现自然变异植株的概率是 0.0045, 试用二项分布计算:

(1) 调查 100 株, 获得两株或两株以上变异植株的概率是多少?

(2) 期望有 0.99 的概率获得 1 株或 1 株以上变异植株, 至少应该调查多少株?

例 3.8: 某小麦品种在田间出现自然变异植株的概率是 0.0045, 试用泊松分布计算:

(1) 调查 100 株, 获得两株或两株以上变异植株的概率是多少?

(2) 期望有 0.99 的概率获得 1 株或 1 株以上变异植株, 至少应该调查多少株?

例 3.9: 设 u 服从正态分布 $N(0,1)$, 试求:

(1) $P(u \leq 1)$

(2) $P(u > 1)$

(3) $P(-2 < u \leq 1.5)$

(4) $P(|u| > 2.58)$

实战 3-u 检验 t 检验

例 4.2 已知某种玉米平均穗重 $\mu_0=300(g)$,标准差 $9.5(g)$.喷施某种生长调节剂后随机抽取 9 个果穗,重量为 308,305,311,298,315,300,321,294,320。问这种调节剂对果穗重量是否有显著影响?

例 4.4 某鱼塘水中的含氧量多年平均为 $4.5 (mg/L)$.现在该鱼塘设 10 个点采集水样,测定水中含氧量分别为 4.33,4.62,3.89,4.14,4.78,4.64,4.52,4.55,4.48,4.26。试检验该次抽样测定的水中含氧量与多年平均值有无显著差别。

例 4.6 现用甲、乙两种发酵法生产青霉素,其产品效率的方差分别为 $var_1=0.46$, $var_2=0.37$ 。用甲方法测得 25 个数据, $aver_1=3.71$; 乙方法测得 30 个数据, $aver_2=3.46$ 。问甲、乙两种方法的收率是否相同?

例 4.8 用高蛋白和低蛋白两种饲料饲养 1 月龄大白鼠,在 3 个月时,测定两组大白鼠的增重量 (g),两组的数据分别为

高蛋白组	134	146	106	119	124	161	107	83	113	129	97	123
低蛋白组	70	118	101	85	107	132	94					

试问两种饲料饲养的大白鼠增重量是否有差别?

课外例题: 某研究人员测得正常细胞 (样本容量为 6) 和肿瘤细胞 (样本容量为 8) 中 p53 基因的表达水平如下,用 t-检验方法分析 p53 基因的表达水平在正常细胞和肿瘤细胞间有无显著差异。

正常	5	6	8	7	6.1	5.5		
肿瘤	4	6.2	5	3	5	4.5	5.1	5.5

实战 4-样本频率的检验

例 1：以前进行细胞培养时，平均每 μL 污染的杂菌为 1.5 个。先用一种新的灭菌方法，在 $1\mu\text{L}$ 中发现 4 个杂菌。是否可据此认为新方法的污染率有变化？

分析：假定每 μL 样品中有 N 个细胞，其中 1.5 个杂菌细胞。杂菌出现的概率（即总体频率 $p=1.5/N$ ）很小，杂菌出现的数量服从泊松分布。因此，检验新方法对应的样本频率 $p_1=4/N$ 是否与老方法总体频率 $p=1.5/N$ 时，应该用泊松分布检验。

例 2：调查包头市居民吃面食或吃米饭的偏好。在包头大街随机调查 10 人，有 8 人喜欢吃面食，2 人喜欢吃米饭。是否可以据此认为包头居民吃面食和米饭的比例不是 1: 1？

例 4.13 某养鸡场规定种蛋的孵化率 $p_0 > 0.80$ 为合格。现对一批种蛋随机抽取 100 枚进行孵化检验，结果有 78 枚孵出。问这批种蛋是否合格？

课外例题：在甲地有 193 人，该病患病率 0.0725；在乙地有 4177 人，某种病患病率是 0.03。两地患病率是否有差异？

实战 5-卡方检验和方差分析

例题 1: 下列表列出对某种药的试验结果，问给药方式对药效是否有影响？

给药方式	有效	无效	总数
口服	58	40	98
注射	64	31	95
总数	122	71	193

例题 2: 某研究人员测得正常细胞（样本容量为 6）和肿瘤细胞（样本容量为 8）中 p53 基因的表达水平如下，用方差分析检验 p53 基因的表达水平在正常细胞和肿瘤细胞间有无显著差异。

正常	5	6	8	7	6.1	5.5		
肿瘤	4	6.2	5	3	5	4.5	5.1	5.5

例 5.1: 鲤鱼遗传试验，以荷包红鲤（红色，隐性）与湘江野鲤（青灰色，显性）杂交，其 F₂ 代获得如下表所列体色分离尾数，问这一资料的实际观察值是否符合孟德尔一对等位基因的遗传规律，即鲤鱼体色青：红=3：1？

体色	青灰色	红色	总和
F ₂ 代观测尾数	1503	99	1602

例 5.4: 现随机抽样对吸烟人群和不吸烟人群是否患有气管炎进行了调查，调查结果如下表。试检验吸烟与患气管炎有无关联。

不同人群	患病	不患病	总和	患病率%
吸烟人群	50	250	300	16.67
不吸烟人群	5	195	200	2.5
总和	55	445	500	

习题 5.3: 某林场狩猎得到 143 只野兔, 其中雄性 57 只, 雌性 86 只。试检验野兔的性别比例是否符合 1: 1?

习题 5.7: 五个小麦品种感染赤霉病的情况如下表。试分析小麦赤霉病的发生是否与品种有关。

品种	A	B	C	D	E	总和
健株数	442	460	478	376	494	2250
病株数	78	39	35	298	50	500
总和	520	499	513	674	544	2750

习题 6.4: 为研究氟对种子发芽的影响, 分别用 0 (对照), 10, 50, 100 (ug/g) 四种不同浓度的氟化钠溶液处理种子, 每种浓度处理的种子用培养皿进行发芽实验 (每皿 50 粒, 每处理重复 3 次), 观察他们的发芽情况, 结果如下表。试作方差分析, 并用 LSD 法和 SSR 法分别进行多重比较。

处理 (ug/g)	芽长		
	重复 1 (cm)	重复 2 (cm)	重复 3 (cm)
0 (对照)	8.9	8.4	8.6
10	8.2	7.9	7.5
50	7	5.5	6.1
100	5	6.3	4.1

实战 6-线性回归与相关分析

例 7.1: 有人研究黏虫孵化历期平均温度与历期天数之间的关系, 试验资料如下。
试建立直线回归方程。

平均温度(°C)	11.8	14.7	15.6	16.8	17.1	18.8	19.5	20.4
历期天数(天)	30.1	17.3	16.7	13.6	11.9	10.7	8.3	6.7

例 7.8: 有人研究黏虫孵化历期平均温度与历期天数之间的关系, 试验资料如下。
试分析历期温度与历期天数之间的相关性。

平均温度(°C)	11.8	14.7	15.6	16.8	17.1	18.8	19.5	20.4
历期天数(天)	30.1	17.3	16.7	13.6	11.9	10.7	8.3	6.7

习题 7.4: 下面是某地区 4 月下旬平均气温与 5 月上旬 50 株棉苗蚜虫头数的资料。

(1) 建立直线回归方程

(2) 对回归系数作假设检验

(3) 该地区 4 月下旬平均气温为 18 摄氏度时, 5 月上旬 50 株棉苗蚜虫预期为多少头?

年份	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
4 月下旬平均气温 (°C)	19.3	26.6	18.1	17.4	17.5	16.9	16.9	19.1	17.9	17.5	18.1	19
5 月上旬蚜虫头数	86	197	8	29	28	29	23	12	14	64	50	112

实战 7-作图(箱线图-柱状图-折线图-线性回归直线图)

1. 用 R 语言读入文本文件 (file1_GC.txt), 第一列数据标记为 “control”, 第二列数据标记为 “treatment”。

- (1) 计算每一列数据的平均值和标准偏差;
- (2) 计算两列数据的平均数之间是否有显著差异 (用 t 检验);
- (3) 画两列数据的箱线图(boxplot);
- (4) 画两列数据的柱状图(barplot)。

例 7.1: 有人研究黏虫孵化历期平均温度与历期天数之间的关系, 试验资料如下。

平均温度(°C)	11.8	14.7	15.6	16.8	17.1	18.8	19.5	20.4
历期天数(天)	30.1	17.3	16.7	13.6	11.9	10.7	8.3	6.7

- (1) 试建立直线回归方程;
- (2) 作散点图并添加回归直线;
- (3) 作历期天数随平均温度变化的折线图。

例 7.8: 有人研究黏虫孵化历期平均温度与历期天数之间的关系, 试验资料如下。

平均温度(°C)	11.8	14.7	15.6	16.8	17.1	18.8	19.5	20.4
历期天数(天)	30.1	17.3	16.7	13.6	11.9	10.7	8.3	6.7

- (1) 试分析历期温度与历期天数之间的相关性。
- (2) 作散点图, 并添加回归直线和相关系数和 p-value