马怀良 金志民 著

基于Excel的生物试验数据分析

冶金工业出版社
www.cnmip.com.cn

基于 Excel 的
生物试验数据分析

马怀良　金志民　著

北　京
冶 金 工 业 出 版 社
2019

内 容 提 要

本书系统地介绍了应用 Microsoft Excel 软件进行生物试验设计和数据分析的方法和经验，内容涵盖了数据整理、描述性统计、一个样本的假设检验、两个样本的假设检验、卡方检验、多个样本的假设检验（方差分析）、回归与相关、协方差分析、随机区组设计统计分析、裂区设计统计分析、拉丁方设计与统计分析、正交设计统计分析 12 类 Excel 数据分析方法。

本书分类系统，内容全面、翔实，适用性强，可供生物学科技工作者阅读，也可供大专院校有关师生参考。

图书在版编目(CIP)数据

基于 Excel 的生物试验数据分析/马怀良，金志民著. —北京：冶金工业出版社，2019.1

ISBN 978-7-5024-7987-9

Ⅰ.①基… Ⅱ.①马… ②金… Ⅲ.①表处理软件—应用—生物学—实验数据—数据处理 Ⅳ.①Q-39

中国版本图书馆 CIP 数据核字(2019)第 006664 号

出 版 人 谭学余

地 址 北京市东城区嵩祝院北巷 39 号 邮编 100009 电话 (010)64027926

网 址 www.cnmip.com.cn 电子信箱 yjcbs@cnmip.com.cn

责任编辑 贾怡雯 美术编辑 郑小利 版式设计 禹 蕊

责任校对 郑 娟 责任印制 李玉山

ISBN 978-7-5024-7987-9

冶金工业出版社出版发行；各地新华书店经销；三河市双峰印刷装订有限公司印刷

2019 年 1 月第 1 版，2019 年 1 月第 1 次印刷

169mm×239mm；14 印张；270 千字；213 页

65.00 元

冶金工业出版社 投稿电话 (010)64027932 投稿信箱 tougao@cnmip.com.cn

冶金工业出版社营销中心 电话 (010)64044283 传真 (010)64027893

冶金工业出版社天猫旗舰店 yjgycbs.tmall.com

前　言

合理的试验设计、严谨的误差控制手段、科学正确的数据统计分析方法，可提高科研成果或结论的准确性和可靠性，是生物学相关专业的教师、科研人员、学生必须掌握的一项技能。但数据统计分析公式多，运算量大、繁琐，笔算或用计算器辅助运算费时、费力，容易出错。最理想的是应用 SPSS、SAS、Minitab、DPS 等专业统计软件。但这些专业统计软件的市场普及率非常低，能熟练使用的人也不多，主要原因有：一是价格昂贵，一般单位和个人财力有限，购买的可能性低；二是对使用者自身的水平有一定的要求，如熟悉编程方法及英文专业术语；三是有的统计软件更新滞后于计算机软硬件配置和操作系统，经常无法安装。

Microsoft Excel 是微软公司开发的办公软件 Microsoft Office 组件之一，提供了 10 大类数百个内置函数，应用非常广泛。尽管 Excel 不是专业统计软件，其统计功能也无法与专业统计软件媲美，但在科研数据处理中具有相当大的优势：Excel 普及率很高，是人们最常用的办公软件；各高校开设 Office 相关课程或培训，师生熟悉 Excel 的操作；Excel 有多种版本，能很好地适应计算机配置和操作系统；Excel 运算的精度通常高于专业统计软件；Excel 支持编制程序，方便应用。

目前，市面上专门介绍生物学科研相关的 Excel 统计方法和技巧的图书较少。为此，作者将应用 Microsoft Excel 软件进行试验设计和数据分析的多年经验加以整理和完善，撰写了本书。数据整理与描述性统计、一个与两个样本的假设检验、附表、函数及用法由金志民执笔，其余由马怀良执笔并负责全书修改与定稿。

本书的特色为：

(1) 知识结构合理。按照国内主流生物统计学教材的结构编写，方便读者应用。

(2) 知识覆盖面广。在常规的生物统计的基础上，增加了数据异常值与正态性检验、符合二项分布的小样本检验（精确法）和非参数检验的 Excel 统计和分析方法。

(3) 自成系统性。在编写顺序上，先简要地列出数据处理分析的原理和安排范例，然后介绍应用 Excel 最简单、快捷的统计分析方法，可配合生物统计学教材使用，也可直接使用。

(4) 应用条件明晰。明确指出各种统计分析方法的应用条件，如单因素方差分析要求各组数据无异常值、服从正态分布和方差同质(等方差)，以防读者误用或错用统计方法。

(5) 适用性广泛。本书适用于 Microsoft Excel 2000 及之后的版本(各版本的一些函数及用法稍有不同，参见本书相关函数及用法)，是从事生物学相关工作的教师、科研人员、学生的实用工具书。

本书能够顺利出版，得益于牡丹江师范学院 2018 年学术专著出版基金项目资助，在此表示衷心的感谢。

由于作者水平所限，书中不足之处，恳请同行、专家及读者批评指正。

作　者

2018 年 8 月

目　录

1　数据整理与描述性统计 …… 1

1.1　数据整理 …… 1

1.1.1　离散型变量 …… 1

1.1.2　连续型变量 …… 4

1.2　数据异常值检验 …… 4

1.3　数据正态性检验 …… 7

1.4　描述性统计 …… 10

1.4.1　平均数 …… 10

1.4.2　变异数 …… 10

2　一个样本的假设检验 …… 15

2.1　符合正态分布的一个样本假设检验 …… 15

2.1.1　总体方差已知 …… 15

2.1.2　总体方差未知——大样本 …… 15

2.1.3　总体方差未知——小样本 …… 16

2.2　符合二项分布的一个样本假设检验 …… 20

2.2.1　精确法——小样本 …… 20

2.2.2　正态理论法——大样本 …… 20

2.3　一个样本的非参数假设检验 …… 24

2.3.1　精确法——小样本 …… 24

2.3.2　正态理论法——大样本 …… 25

3　两个样本的假设检验 …… 29

3.1　符合正态分布的两个样本假设检验 …… 29

3.1.1　成组数据（总体方差已知） …… 29

3.1.2　成组数据（总体方差未知）——大样本 …… 32

3.1.3　成组数据（总体方差未知）——小样本 …… 34

3.1.4　配对数据 …… 40

3.2 符合二项分布的两个样本假设检验 …… 43
3.2.1 精确法——小样本 …… 43
3.2.2 正态理论法——大样本 …… 44
3.3 两个样本的非参数假设检验 …… 47
3.3.1 配对数据 …… 47
3.3.2 成组数据 …… 52

4 卡方检验 …… 57

4.1 假设检验原理 …… 57
4.2 适合性检验 …… 57
4.3 独立性检验 …… 60
4.3.1 2×2 列联表 …… 60
4.3.2 *r*×*c* 列联表 …… 61

5 多个样本的假设检验 …… 65

5.1 符合正态分布的多个样本假设检验 …… 65
5.1.1 单因素方差分析 …… 65
5.1.2 双因素方差分析——无重复值 …… 75
5.1.3 双因素方差分析——有重复值 …… 79
5.2 多个样本的非参数检验法 …… 83
5.2.1 kruskal-wallis 秩和检验 …… 84
5.2.2 kruskal-wallis 两两比较 …… 84

6 回归与相关 …… 94

6.1 一元线性回归与相关分析 …… 94
6.1.1 一元线性回归分析 …… 94
6.1.2 一元线性相关分析 …… 95
6.2 可直线化的一元曲线回归 …… 99
6.2.1 指数函数曲线 …… 100
6.2.2 对数函数曲线 …… 102
6.2.3 幂函数曲线 …… 104
6.2.4 倒数函数曲线 …… 104
6.2.5 Logistic 生长曲线 …… 106
6.3 多元线性回归 …… 107
6.4 多项式回归 …… 114

6.4.1　一元多项式回归 …… 114
6.4.2　多元多项式回归 …… 117
6.5　规划求解 …… 117

7　协方差分析 …… 123

7.1　方差齐性检验 …… 123
7.2　检验各组回归直线是否平行 …… 123
7.3　公共回归系数的显著性检验 …… 124
7.4　协方差分析 …… 124
7.5　多重比较 …… 125

8　常用试验设计及统计分析 …… 132

8.1　随机区组设计及统计分析 …… 132
8.1.1　单因素随机区组设计及统计分析 …… 132
8.1.2　双因素随机区组设计及统计分析 …… 133
8.2　裂区设计及统计分析 …… 140
8.3　拉丁方设计及统计分析 …… 147
8.3.1　拉丁方设计 …… 147
8.3.2　拉丁方设计统计分析 …… 151
8.4　正交设计及统计分析 …… 154
8.4.1　随机设计的正交统计分析 …… 154
8.4.2　随机区组设计的正交统计分析 …… 163

附录 …… 169

附录1　相关表格 …… 169
附录2　相关函数及用法 …… 201

参考文献 …… 213

1 数据整理与描述性统计

本章数字资源

1.1 数据整理

一般情况下，样本容量 $n \leqslant 30$ 时不必分组整理数据。但 $n>30$ 时，可将数据整理成若干组，制成次数分布表或次数分布图，以便直观地反映试验数据的特征。

1.1.1 离散型变量

1.1.1.1 单项分组法

样本观测值波动范围较小时，以样本观测值为分组依据。

【例 1-1】 随机采 100 个麦穗，对每穗小穗数进行计数，结果如图 1-1 所示，请制作次数分布表。

(1) 应用 COUNTIF 函数。

数据格式化如图 1-1 所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	小穗数											小穗数 (x)	次数 (f)	频率
2	18	15	17	19	16	15	20	18	19	17		15		
3	17	18	17	16	18	20	19	17	16	18		16		
4	17	16	17	19	18	18	17	17	17	18		17		
5	18	15	16	18	18	18	17	20	19	18		18		
6	17	19	15	17	17	17	16	17	18	18		19		
7	17	19	19	17	19	17	18	16	18	17		20		
8	17	19	16	16	17	17	17	16	17	16		合计		
9	18	19	18	18	19	19	20	15	16	19				
10	18	17	18	20	19	17	18	17	17	16				
11	15	16	18	17	18	16	17	19	19	17				

图 1-1 数据格式化

在 M2 单元格输入“=COUNTIF($A$2:$J$11,L2)”，回车；拖动 M2 单元格填充柄至 M7 单元格。

在 M8 单元格输入“=SUM(M2:M7)”，回车；拖动 M8 单元格填充柄至 N8 单元格。

在 N2 单元格输入“=M2/$M$8”，回车；拖动 N2 单元格填充柄至 N7 单元格。

结果如图 1-2 所示。

	L	M	N
1	小穗数（x）	次数（f）	频率
2	15	6	0.06
3	16	15	0.15
4	17	32	0.32
5	18	25	0.25
6	19	17	0.17
7	20	5	0.05
8	合计	100	1.00

图 1-2 统计结果

（2）应用 FREQUENCY 函数。

选取 M2:M7 区域，直接输入“=”，此时等号自动出现在 M2 单元格，然后输入完整函数“=frequency(A2:J11,L2:L7)”，如图 1-3 所示。

	K	L	M	N
1		小穗数（x）	次数（f）	频率
2		=frequency(A2:J11,L2:L7)		
3		16		#DIV/0!
4		17		#DIV/0!
5		18		#DIV/0!
6		19		#DIV/0!
7		20		#DIV/0!
8		合计	0	#DIV/0!

图 1-3 输入完整函数

最后同时按下“Ctrl+Shift+Enter”三个键，完成输入，结果如图 1-4 所示。其余步骤与（1）相同。

	K	L	M	N
1		小穗数（x）	次数（f）	频率
2		15	6	0.06
3		16	15	0.15
4		17	32	0.32
5		18	25	0.25
6		19	17	0.17
7		20	5	0.05
8		合计	100	1

图 1-4 统计结果

1.1.1.2 区间分组法

样本观测值波动范围较大，可采用一定范围的区间为一组，即区间分组法。

【例 1-2】 某鸡场调查获得 100 只鸡年产蛋数，请制作次数分布表。

数据格式化如图 1-5 所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	产蛋数											产蛋数(x)			次数(f)	频率
2	200	208	210	213	216	215	213	218	219	216		200	~	209		
3	220	221	222	223	224	225	226	227	228	229		210	~	219		
4	222	223	224	225	226	230	231	232	233	234		220	~	229		
5	234	235	236	237	238	239	230	231	232	233		230	~	239		
6	230	231	232	233	234	240	241	242	243	244		240	~	249		
7	245	246	247	248	249	241	242	243	244	245		250	~	259		
8	240	241	242	243	244	245	246	247	250	251		260	~	269		
9	252	253	254	255	256	250	251	252	253	254		270	~	279		
10	255	256	257	258	259	262	263	264	265	266		280	~	289		
11	267	268	269	270	270	276	274	285	287	296		290	~	299		
12												合 计				

图 1-5 数据格式化

选取 O2:O11 区域，直接输入“=frequency(A2:J11,N2:N11)”，然后同时按下“Ctrl+Shift+Enter”三个键，完成输入（方法见图 1-3 和图 1-4）。

在 O12 单元格输入“=SUM(O2:O11)”，回车；拖动 O12 单元格填充柄至 P12 单元格。

在 P2 单元格中输入“=O2/$O$12”，回车；拖动 P2 单元格填充柄至 P11 单元格。

结果如图 1-6 所示。

	L	M	N	O	P
1	产蛋数(x)			次数(f)	频率
2	200	~	209	2	0.02
3	210	~	219	8	0.08
4	220	~	229	15	0.15
5	230	~	239	20	0.20
6	240	~	249	23	0.23
7	250	~	259	17	0.17
8	260	~	269	8	0.08
9	270	~	279	4	0.04
10	280	~	289	2	0.02
11	290	~	299	1	0.01
12	合 计			100	1.00

图 1-6 统计结果

对区间分组计数来说，O2 表示小于等于 209 的个数，O3 表示大于 209 但小于等于 219 的个数，O4 表示大于 219 但小于等于 229 的个数，以此类推。

1.1.2 连续型变量

连续型变量采用组距式分组法，即通过确定组数、计算组距和每组组限及组中值进行分组。方法与 1.1.1.2 中的【例 1-2】操作步骤完全相同。

【例 1-3】 对某地 100 例 30~40 岁健康男子血清总胆固醇（mol/L）进行组距式分组，结果如图 1-7 所示，具体操作过程由读者自行完成。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	血清总胆固醇											组别		组中值	次数(f)	频率
2	4.77	3.37	6.14	3.95	3.56	4.23	4.31	4.71	5.69	4.12		2.5~	3.0	2.75	1	0.01
3	4.56	4.37	5.39	6.30	5.21	7.22	5.54	3.93	5.21	6.51		3.0~	3.5	3.25	8	0.08
4	5.18	5.77	4.79	5.12	5.20	5.10	4.70	4.74	3.50	4.69		3.5~	4.0	3.75	8	0.08
5	4.38	4.89	6.25	5.32	4.50	4.63	3.61	4.44	4.43	4.25		4.0~	4.5	4.25	24	0.24
6	4.03	5.85	4.09	3.35	4.08	4.79	5.30	4.97	3.18	3.97		4.5~	5.0	4.75	24	0.24
7	5.16	5.10	5.85	4.79	5.34	4.24	4.32	4.77	6.36	6.38		5.0~	5.5	5.25	17	0.17
8	4.88	5.55	3.04	4.55	3.35	4.87	4.17	5.85	5.16	5.09		5.5~	6.0	5.75	9	0.09
9	4.52	4.38	4.31	4.58	5.72	6.55	4.76	4.61	4.17	4.03		6.0~	6.5	6.25	6	0.06
10	4.47	3.40	3.91	2.70	4.60	4.09	5.96	5.48	4.40	4.55		6.5~	7.0	6.75	2	0.02
11	5.38	3.89	4.60	4.47	3.64	4.34	5.18	6.14	3.24	4.90		7.0~	7.5	7.25	1	0.01
12												合 计			100	1.00

图 1-7 血清总胆固醇数据

1.2 数据异常值检验

异常值是指在正态分布的样本中，明显偏离其余观测值的个别观测值，通常为高端值或（和）低端值。异常值的存在影响统计分析结果的可靠性、准确性和精确性。

狄克逊法适用于总体标准差未知，样本容量或重复数 n 在 3~100 样本数据。狄克逊法有单侧检验和双侧检验两种，前者只对疑似异常值的最大值或最小值进行检验，判断是否存在异常值；而后者同时对样本中的最大值与最小值进行检验结果进行比较，然后判断最大值或最小值是否为异常值。因此，双侧检验对科研数据更为合适和合理，方法为：

（1）将各数据从小到大排列，依次记为 x_1、x_2、x_3、…、x_{n-1}、x_n。

（2）分别计算代表最大值 x_n 和最小值 x_1 的统计量 D_n 和 D'_n，见表 1-1。

表 1-1 异常值检验统计量

n	检验最大值	检验最小值
3~7	$D_n = \dfrac{x_n - x_{n-1}}{x_n - x_1}$	$D'_n = \dfrac{x_2 - x_1}{x_n - x_1}$

续表 1-1

n	检验最大值	检验最小值
8~10	$D_n=\frac{x_n-x_{n-1}}{x_n-x_2}$	$D'_n=\frac{x_2-x_1}{x_{n-1}-x_1}$
11 ~ 13	$D_n=\frac{x_n-x_{n-2}}{x_n-x_2}$	$D'_n=\frac{x_3-x_1}{x_{n-1}-x_1}$
14 ~ 100	$D_n=\frac{x_n-x_{n-2}}{x_n-x_3}$	$D'_n=\frac{x_3-x_1}{x_{n-2}-x_1}$

(3) 查附表 1，临界值 $D_{0.95}$（双侧）。

(4) 判断异常值，当 $D_n>D'_n$ 且 $D_n>D_{0.95}$ 时，判定最大值 x_n 为异常值；当 $D'_n>D_n$ 且 $D'_n>D_{0.95}$ 时，判定最小值 x_1 为异常值；不符合以上 2 个条件的，判定没有异常值。

(5) 复检，剔除异常值，然后再重复 (1)~(4) 步，直到没有异常值为止。值得注意的是复检时，n 和 $D_{0.95}$ 会发生变化，D_n 和 D'_n 的计算方法也可能会变。

【例 1-4】 测定某纤维素酶的活力（IU/g），结果为 6.0、5.8、6.1、5.5、6.6、5.4、9.1、3.3、5.3、6.9，判断该组数据有无异常值.

将数据录入 A 列，临界值 $D_{0.95}$ 录入 G:H 列，并对异常值检验进行格式化，如图 1-8 所示。

	A	B	C	D	E	F	G	H
1	数据		异常值检验				n	$D_{0.95}$
2	6.0		条件	总体方差未知, $3\leqslant n\leqslant 100$			3	0.970
3	5.8		方法	Dixon检验法(双侧)			4	0.829
4	6.1	i	高端值		低端值		5	0.710
5	5.5	1	第1大值		第1小值		6	0.628
6	6.6	2	第2大值		第2小值		7	0.569
7	5.4	3	第3大值		第3小值		8	0.608
8	9.1						9	0.564
9	3.3		n		$D_{0.95}$		10	0.530
10	5.3		D		D'		11	0.619
11	6.9		结论				12	0.583

图 1-8 数据格式化

在 D5 单元格输入“=LARGE(A:A,B5)”，回车；拖动 D5 单元格填充柄至 D7 单元格。

在 F5 单元格输入“=SMALL(A:A,B5)”，回车；拖动 F5 单元格填充柄至 F7 单元格。

在 D9 单元格输入“=COUNT(A:A)”，回车。

在 F9 单元格输入“=LOOKUP(D9,G:H)”，回车。

在 D10 单元格输入“=IF(D9>13,(D5-D7)/(D5-F7),IF(D9>10,(D5-D7)/(D5-F6),IF(D9>7,(D5-D6)/(D5-F6),(D5-D6)/(D5-F5))))”，回车。

在 F10 单元格输入“=IF(D9>13,(F7-F5)/(D7-F5),IF(D9>10,(F7-F5)/(D6-F5),IF(D9>7,(F6-F5)/(D6-F5),(F6-F5)/(D5-F5))))”，回车。

在 D11 单元格输入“=IF(AND(D10>F10,D10>F9),"异常值是:"&C5&D5,IF(AND(F10>D10,F10>F9),"异常值是:"&E5&F5,"没有异常值"))”，回车。

结果如图 1-9 所示。

	A	B	C	D	E	F	G	H
1	数据		异常值检验				n	$D_{0.95}$
2	6.0		条件	总体方差未知，$3 \leqslant n \leqslant 100$			3	0.970
3	5.8		方法	Dixon检验法(双侧)			4	0.829
4	6.1	i	高端值		低端值		5	0.710
5	5.5	1	第1大值	9.1	第1小值	3.3	6	0.628
6	6.6	2	第2大值	6.9	第2小值	5.3	7	0.569
7	5.4	3	第3大值	6.6	第3小值	5.4	8	0.608
8	9.1						9	0.564
9	3.3		n	10	$D_{0.95}$	0.530	10	0.530
10	5.3		D	0.5789	D'	0.5556	11	0.619
11	6.9		结论	异常值是:第1大值9.1			12	0.583

图 1-9 统计结果

图 1-9 表明，数据 9.1 是异常值。直接在 A 列删除 9.1，结果如图 1-10 所示。

图 1-10 表明，数据 3.3 是异常值。直接在 A 列删除 3.3，结果如图 1-11 所示。

图 1-11 表明，删除异常值 9.1 和 3.3 后，剩余的 8 个数据就没有异常值，异常值检验到此结束。

本程序可自动进行异常值检验，数据无须排序；只需要按结论的提示删除 A 列中的相应数据即可；为防止临界值被误删或修改，可将 G:H 冻结或隐藏；如担心程序被误删或修改，也可冻结 B:F(G:H 隐藏的情况) 或 B:H 冻结。如不想录入临界值 G:H 列，删除 F9 单元格内的公式，根据 n 查附表 1 中的 $D_{0.95}$

	A	B	C	D	E	F	G	H
1	数据		异常值检验				n	$D_{0.95}$
2	6.0		条件	总体方差未知, $3 \leqslant n \leqslant 100$			3	0.970
3	5.8		方法	Dixon检验法(双侧)			4	0.829
4	6.1	i	高端值		低端值		5	0.710
5	5.5	1	第1大值	6.9	第1小值	3.3	6	0.628
6	6.6	2	第2大值	6.6	第2小值	5.3	7	0.569
7	5.4	3	第3大值	6.1	第3小值	5.4	8	0.608
8							9	0.564
9	3.3		n	9	$D_{0.95}$	0.564	10	0.530
10	5.3		D	0.1875	D'	0.6061	11	0.619
11	6.9		结论	异常值是:第1小值3.3			12	0.583

图 1-10 删除异常值“9.1”

	A	B	C	D	E	F	G	H
1	数据		异常值检验				n	$D_{0.95}$
2	6.0		条件	总体方差未知, $3 \leqslant n \leqslant 100$			3	0.970
3	5.8		方法	Dixon检验法(双侧)			4	0.829
4	6.1	i	高端值		低端值		5	0.710
5	5.5	1	第1大值	6.9	第1小值	5.3	6	0.628
6	6.6	2	第2大值	6.6	第2小值	5.4	7	0.569
7	5.4	3	第3大值	6.1	第3小值	5.5	8	0.608
8							9	0.564
9			n	8	$D_{0.95}$	0.608	10	0.530
10	5.3		D	0.2000	D'	0.0769	11	0.619
11	6.9		结论	没有异常值			12	0.583

图 1-11 删除异常值“3.3”

直接输入即可。

1.3 数据正态性检验

科研数据的正态性检验常采用夏皮洛-威尔克（Shapiro-Wilk）检验法，适用于 $8 \leqslant n \leqslant 50$ 的样本，但对 $n<8$ 且偏离正态分布的小样本效果较差，方法为：

（1）将各数据从小到大排列，依次记为 x_1、x_2、x_3、…、x_{n-1}、x_n。

（2）计算 K，当 n 为偶数时，$K=n/2$；当 n 为奇数时，$K=(n-1)/2$。

（3）查附表 2 中 $\alpha_i(i=1, 2, \cdots, K)$ 的值。

（4）计算统计量 W。

$$L = \sum_{i=1}^{K} \alpha_i (x_{n+1-i} - x_i)$$

$$SS = \sum (x - \bar{x})^2 = \sum x^2 - (\sum x)^2/n$$

$$W = \frac{L^2}{SS}$$

（5）查附表 3 中的临界值 $W_{0.01}$ 或 $W_{0.05}$。

（6）作结论，$W \leqslant W_{0.05}$，该组数据在 0.05 水平上不服从正态分布；$W > W_{0.05}$，该组数据在 0.05 水平上服从或符合正态分布。读者根据实际情况选择合适的显著水平（0.05 或 0.01，一般选用 0.05）。

【例 1-5】 检验【例 1-4】中剔除异常值后的数据是否符合正态分布。

将数据录入 A 列，附表 3 中 n 在 3~50 的 $W_{0.05}$ 录入 J:K 列；附表 2 中 n 在 3~50、k 在 1~25 的 α_i 录入 M:AL 列；并对异常值检验进行格式化，如图 1-12 所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	数据	i	$x_{n+1-i}-x_i$	α_i		正态性检验				n	$W_{0.05}$			k			
2	6.0	1				条件	3≤n≤50			3	0.767		n	1	2	3	4
3	5.8	2				方法	Shapiro-Wilk检验法			4	0.748		2	0.7071			
4	6.1	3								5	0.762		3	0.7071			
5	5.5	4				n		W		6	0.788		4	0.6872	0.1677		
6	6.6	5				k		$W_{0.05}$		7	0.803		5	0.6646	0.2413		
7	5.4	6								8	0.818		6	0.6431	0.2806	0.0875	
8	5.3	7				结论				9	0.829		7	0.6233	0.3031	0.1401	
9	6.9	8								10	0.842		8	0.6052	0.3164	0.1743	0.0561
10		9								11	0.850		9	0.5888	0.3244	0.1976	0.0947
11		10								12	0.859		10	0.5739	0.3291	0.2141	0.1224
12		11								13	0.866		11	0.5601	0.3315	0.226	0.1429
13		12								14	0.874		12	0.5475	0.3325	0.2347	0.1586
14		13								15	0.881		13	0.5359	0.3325	0.2412	0.1707
15		14								16	0.887		14	0.5251	0.3318	0.246	0.1802
16		15								17	0.892		15	0.515	0.3306	0.2495	0.1878
17		16								18	0.897		16	0.5056	0.329	0.2521	0.1939
18		17								19	0.901		17	0.4968	0.3273	0.254	0.1988
19		18								20	0.905		18	0.4886	0.3253	0.2553	0.2027
20		19								21	0.908		19	0.4808	0.3232	0.2561	0.2059
21		20								22	0.911		20	0.4734	0.3211	0.2565	0.2085
22		21								23	0.914		21	4643	0.3185	0.2578	0.2119
23		22								24	0.916		22	0.459	0.3156	0.2571	0.2131
24		23								25	0.918		23	0.4542	0.3126	0.2563	0.2139
25		24								26	0.920		24	0.4493	0.3098	0.2554	0.2145
26		25								27	0.923		25	0.445	0.3069	0.2543	0.2148

图 1-12 数据格式化

在 C2 单元格输入“=IF(B2>$G$6,0,LARGE(A:A,B2)-SMALL(A:A,B2))”，回车；拖动 C2 单元格填充柄至 C26 单元格。

在 D2 单元格输入“=IF(B2>$G$6,0,INDEX($M$2:$AL$51,$G$5,B2+1))”，回车；拖动 D2 单元格填充柄至 D26 单元格。

在 G5 单元格输入“=COUNT(A:A)”，回车。

在 G6 单元格输入“=IF(ISEVEN(G5),G5/2,(G5-1)/2)”，回车。

在 I5 单元格输入“=SUMPRODUCT(C:C,D:D)^2/DEVSQ(A:A)”，回车。

在 I6 单元格输入“=LOOKUP(G5,J:K)”，回车。

在 G8 单元格输入“=IF(I5>I6,"服从正态分布","不服从正态分布")”，回车。

结果如图 1-13 所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	数据	i	$x_{n+1-i}-x_i$	α_i		正态性检验				n	$W_{0.05}$			k			
2	6.0	1	1.6	0.6052		条件	$3\leqslant n\leqslant 50$			3	0.767		n	1	2	3	4
3	5.8	2	1.2	0.3164		方法	Shapiro-Wilk检验法			4	0.748		2	0.7071			
4	6.1	3	0.6	0.1743						5	0.762		3	0.7071			
5	5.5	4	0.2	0.0561		n	8	W	0.9316	6	0.788		4	0.6872	0.1677		
6	6.6	5	0.0	0		k	4	$W_{0.05}$	0.818	7	0.803		5	0.6646	0.2413		
7	5.4	6	0.0	0						8	0.818		6	0.6431	0.2806	0.0875	
8	5.3	7	0.0	0		结论	服从正态分布			9	0.829		7	0.6233	0.3031	0.1401	
9	6.9	8	0.0	0						10	0.842		8	0.6052	0.3164	0.1743	0.0561
10		9	0.0	0						11	0.850		9	0.5888	0.3244	0.1976	0.0947
11		10	0.0	0						12	0.859		10	0.5739	0.3291	0.2141	0.1224
12		11	0.0	0						13	0.866		11	0.5601	0.3315	0.226	0.1429
13		12	0.0	0						14	0.874		12	0.5475	0.3325	0.2347	0.1586
14		13	0.0	0						15	0.881		13	0.5359	0.3325	0.2412	0.1707
15		14	0.0	0						16	0.887		14	0.5251	0.3318	0.246	0.1802
16		15	0.0	0						17	0.892		15	0.515	0.3306	0.2495	0.1878
17		16	0.0	0						18	0.897		16	0.5056	0.329	0.2521	0.1939
18		17	0.0	0						19	0.901		17	0.4968	0.3273	0.254	0.1988
19		18	0.0	0						20	0.905		18	0.4886	0.3253	0.2553	0.2027
20		19	0.0	0						21	0.908		19	0.4808	0.3232	0.2561	0.2059
21		20	0.0	0						22	0.911		20	0.4734	0.3211	0.2565	0.2085
22		21	0.0	0						23	0.914		21	4643	0.3185	0.2578	0.2119
23		22	0.0	0						24	0.916		22	0.459	0.3156	0.2571	0.2131
24		23	0.0	0						25	0.918		23	0.4542	0.3126	0.2563	0.2139
25		24	0.0	0						26	0.920		24	0.4493	0.3098	0.2554	0.2145
26		25	0.0	0						27	0.923		25	0.445	0.3069	0.2543	0.2148

图 1-13 统计结果

图 1-13 表明，删除异常值 9.1 和 3.3 后，剩余的 8 个数据服从正态分布。

本程序能自动进行正态性检验，数据无须排序；为防止临界值被误删或修改，可将 J:AL 冻结或隐藏。如不想录入 J:AL 列，可删除 D2:D26 和 I6 单元格内的公式，根据 n 查附表 2 和附表 3 中的 α_i 和 $W_{0.05}$ 直接输入即可。但必须使 D2:D26 无 α_i 值的单元格为零。