

同异反灰色相关分析方法及其在小麦中的应用

郭瑞林

(河南省安阳市农业科学研究所, 河南 安阳 455000)

摘要: 运用数理统计、集对分析和灰色系统理论, 提出了一种新的相关分析方法——同异反灰色相关分析法。介绍了这种分析方法的原理与分析步骤, 指出了相关系数的灰色性, 分析了相关系数中灰色性产生的根源, 解析了相关系数的构成, 从相关系数中分离出灰色部分, 并提出了同异反分析意义下的相关系数灰色度概念, 揭示出数理统计、灰色系统、集对分析三者之间的内在联系。与此同时, 结合实例, 介绍了这种分析方法在小麦中的具体应用。结果表明, 在研究因素之间的相关关系时, 除了考虑相关系数的显著性测验外, 还要考虑相关系数的灰色度问题。简言之, 即只有在 r 既达到显著水平, 又满足较低灰色度 (姑且定为 0.3 以下) 的情况下, 得到的结论才是比较可靠的。在此基础上, 讨论了尚待进一步探讨的问题。表 4, 参 8。

关键词: 同异反; 灰色系统; 相关分析; 小麦

中图分类号: S11+7 **文献标识码:** A **文章编号:** 1001-0068(2005)03-0170-05

Identical-discrepant-contrary grey correlation analysis method and its application in wheat

GUO Rui-lin

(Anyang Institute of Agricultural Sciences, He'nan, Anyang 455000, China)

Abstract: A new correlation analysis method—identical-discrepant-contrary grey correlation analysis method was put forward using mathematical statistics, set pair analysis and grey system theory. The principle and its analysis steps of the method were introduced, the grey nature of r was pointed out, the cause of producing r 's grey nature was analysed, the composition of r was resolved, the grey fraction was segregated from r , and the conception of grey degree of r in the sense of identical-discrepant-contrary analysis was advanced which revealed the internal connection among mathematical statistics, grey system and set pair analysis. At the same time, the concrete application of the method in wheat was introduced by combining with the example. The result show that while studying the correlation relationship among factors, the problem of grey degree of r should be considered besides the significance test of r , that is to say, only in the conditions of r reaching the significance level and satisfying lower grey degree (temporarily established it as below 0.3), the result obtained is reliable. Based on this, the problems remained to be further studied were discussed.

Key words: identical-discrepant-contrary; grey system; correlation analysis; wheat

相关分析是农业科学研究过程中常用的一种数理统计方法^[1,2]。实践表明, 这种方法为农业科学家揭示隐含在纷纭复杂的农业现象背后的规律性提供了一种十分有效的工具和手段, 目前已日臻完善和成熟。然而, 由于农业系统信息不完全导致的灰色性^[3,4], 使得相关分析的结果常常带有某些灰色成分。譬如相关系数 r , 一般地表示自变量与因变量两性状之间线性关系的密切程度, 是 n 个自变量与相应的因变量之间相关关系的总和。其中, 既包括自变量与因变量的正相关关系, 又包括自变量与因变量的负相关关系, 还包括自变量与因变量的零相关关系。而这些信息在相

关系数 r 中是反映不出来的。因此, 我们说相关系数 r 具有一定程度的灰色性。为剖析这种灰色性, 我们在传统相关分析的基础上, 运用集对分析^[5]和灰色分析原理, 提出了同异反灰色相关分析方法, 供从事作物研究工作的同行参考。

1 材料与方 法

1.1 同异反灰色相关分析方法的原理

定义 1: 给定 2 个集合 X (自变量) 和 Y (因变量), 并设这 2 个集合组成集对 $H = (X, Y)$, 在某个具体的问题背景 (设为 W) 下, 我们对集对 H 的特性展开分析, 共得到 N 个特性, 其中:

收稿日期: 2004-05-27.

基金项目: 国家农业科技成果转化资金项目 (02EFN21400396)。

作者简介: 郭瑞林 (1961-), 男, 河南林州人, 研究员, 研究方向为高产、优质、多抗小麦新品种选育和作物灰色育种学的建构及应用。

有 S 个特性为集对 H 中的两个集合 X 和 Y 共同具有; 有 P 个特性在集合 X 和 Y 上相互对立, 其余的 $F = N - S - P$ 个特性在集合 X 和 Y 上既不相互对立, 又不为这 2 个集合所共同具有, 则称比值 S/N 为这 2 个集合在问题 W 下的同一度, 简称同一度; F/N 为这 2 个集合在问题 W 下的差异度, 简称差异度; P/N 为这 2 个集合在问题 W 下的对立度, 简称对立度, 并用公式

$$\mu(W) = S/N + (F/N)i + (P/N)j \quad (1)$$

加以统一的表示。为简便计, 可令 $S/N = a$, $F/N = b$, $P/N = c$, 则 (1) 式可简写成

$$\mu(W) = a + bi + cj \quad (2)$$

式中, $\mu(W)$ 称为 X, Y 两个集合的联系度。 i, j 分别是差异度 b 和对立度 c 的系数。规定 $a, b, c \in [0, 1]$, $i \in [-1, 1]$, $j = -1$ 。而且 a, b, c 满足规一化条件, 即有关系式:

$$a + b + c = 1 \quad (3)$$

一般说来, 同一度 a (简称“同”) 和对立度 c (简称“反”) 是相对确定的, 由于 i 在 -1 与 1 之间视具体情况不确定取值, 所以差异度 (简称“异”) 是相对不确定的。同异反是事物相互联系的一种方法, 由于同与反的相对确定, 异的相对不确定, 因此, 同异反分析又是一种既确定又不确定的分析方法, 因而具有辩证性。

令自变量 $X = (x_1, x_2, \dots, x_n)$, 因变量 $Y = (y_1, y_2, \dots, y_n)$ 。为研究方便起见, 分别将其各自的分量按小、中、大均分为 3 个子区间, 即 $[\min X, \min X + (\max X - \min X) / 3]$ 、 $(\min X + (\max X - \min X) / 3, \min X + 2(\max X - \min X) / 3)$ 、 $(\min X + 2(\max X - \min X) / 3, \max X)$ 和 $[\min Y, \min Y + (\max Y - \min Y) / 3]$ 、 $(\min Y + (\max Y - \min Y) / 3, \min Y + 2(\max Y - \min Y) / 3)$ 、 $(\min Y + 2(\max Y - \min Y) / 3, \max Y)$ 。其中, $\min X$ 和 $\min Y$ 分别表示 X, Y 两个集合中的最小值, $\max X$ 和 $\max Y$ 则分别表示 X, Y 两个集合中的最大值。

分别使自变量 X 和因变量 Y 中各分量的数值与上述 3 个子区间对号入座, 譬如 X 中某分量的数值属于 $[\min X, \min X + (\max X - \min X) / 3]$ 区间, 则标上“小”; 属于 $(\min X + (\max X - \min X) / 3, \min X + 2(\max X - \min X) / 3)$ 区间, 则标上“中”; 否则, 标上“大”。 Y 中各分量与此类同。这样, X, Y 各分量的数值集合就变成了由

“小”、“中”和“大”组成的文字集合。令 [小, 小]、[中, 中]、[大, 大] 3 类子集为“同” (表示正相关趋势); [小, 大]、[大, 小] 2 类子集为“反” (表示负相关趋势); 其余的 [小, 中]、[中, 小]、[中, 大]、[大, 中] 等子集为“异” (表示正相关和负相关趋势不确定)。统计 X, Y 两个集合中同、异、反 3 类子集所占份额, 写成 (2) 式即 $a + bi + cj$ 的形式。这样, 对 a, b, c 三者的关系展开分析即所谓同异反分析。

与此同时, 我们再对集合 X, Y 进行相关分析, 其相关关系及其相关系数 r 的定义与公式参见文献 [1] 和 [2], 不予赘述。

将同异反分析与相关分析稍加比较, 不难发现, 相关分析中的相关系数 r 与同异反分析中的同一度 a 、差异度 b 、对立度 c 有着密切联系。一般地, 同一度 a 越大, 表明两个集合的正相关趋势越强; 对立度 c 越大, 表明两个集合的负相关趋势越强; 差异度 b 则是介于正、负相关趋势两者之间的一种不确定趋势, 在一定条件下 (如补充一定信息), 可以向正、负两个方向发生转化。当 i 取正值时, 可转化为正相关, i 取负值时, 可转化为负相关。这就是相关系数 r 与同一度 a 、差异度 b 、对立度 c 四者之间的辩证关系。

事实上, 由于相关系数 r 中混杂着各自变分量与因变分量的正、负和零相关趋势, 因而具有一定程度的灰色性。也就是说, 相关系数 r 可分解为 $|ra|$ 部分的正相关、 $|rc|$ 部分的负相关和 $|rb|$ 部分的不确定相关。当 r 为正时, 上述 3 部分中, 只有 ra 部分的相关是可以肯定的, 而其余的 $rb + rc$ 部分则是不能肯定的; 而当 r 为负时, 只有 rc 部分的相关是可以肯定的, 其余的 $ra + rb$ 部分则是不能肯定的。不能肯定的部分是由信息不完全造成的, 所以我们称这部分为灰色部分。为此, 需要引入“相关系数的灰色度”的概念。

定义 2: 若自变量 X 与因变量 Y 的同异反联系度为 $\mu(X, Y) = a + bi + cj$, 该自变量与因变量 Y 的相关系数为 r ($r > 0$), 则 $(rb + rc) / r$ 为相关系数在同异反分析意义下的一种灰色度, 并用 $f_G(r)$ 表示, 即有

$$f_G(r) = (rb + rc) / r = b + c = 1 - a \quad (r > 0) \quad (4)$$

当 $r < 0$ 时, 则其相关系数的灰色度为

$$f_G(r) = (ra + rb) / r = a + b = 1 - c \quad (r < 0) \quad (5)$$

从式 (4) 和式 (5) 可知, 当相关系数为正时, 相关系数的灰色度与同一度 a 关系密切, a

越大,表明相关系数中的可肯定部分所占成分越多,其灰色度就越小。反之,当相关系数为负时,相关系数的灰色度与对立度 c 关系密切, c 越大,表明相关系数中可肯定的部分所占份额越多,其灰色度也越小。

由此可见,只有当相关系数既达到显著水平,灰色度又较低(姑且规定 $f_G(r) \leq 0.3$) 的情况下,相关系数 r 所表示的意义才比较真实可信。据此,即可对相关系数 r 的灰色性展开分析,从而决定相关系数 r 是否可以用于指导实践。

1.2 同异反灰色相关分析的分析步骤

1.2.1 确定研究对象 X 和 Y , 并搜集整理拟进行相关分析的试验数据。

1.2.2 采用传统相关分析方法^[1,2], 求出相关系数

r , 并进行显著性测验。

1.2.3 对达到显著水平的相关系数 r 进行同异反灰色相关分析。

按照上述方法对自变量 X 和因变量 Y 的分析数据进行适当处理,并据(4)和(5)式求出相关系数的灰色度,进行同异反灰色相关分析。

1.3 数据来源

分析数据来自河南省安阳市农业科学研究所小麦研究室。该室 1992~1994 年选用 10 个有代表性的品种(系),稀点播(行长 3m,行距 0.3m,株距 0.1m)种植,每品种 2 行区,60 株。收获时,各选取 10 株脱粒考种,对产量性状之间的关系进行研究。各品种产量性状平均观察值列于表 1。

表 1 小麦各品种单株产量与产量三因素观察值

性 状	品种代码									
	1	2	3	4	5	6	7	8	9	10
单株产量 g	19.3	17.8	21.5	14.4	17.7	15.9	16.4	16.9	14.6	14.9
单株穗数	13.0	12.3	13.6	10.2	12.3	11.1	11.4	11.9	8.4	10.9
穗粒数	43.6	35.7	42.3	35.9	40.8	37.3	36.9	36.8	41.7	36.5
千粒重 g	34.1	40.5	37.7	39.8	35.0	38.7	39.0	38.7	41.5	37.3

2 结果与分析

2.1 产量性状之间的相关分析

按照参考文献 [1]、[2] 介绍的统计方法,得到小麦单株产量及其产量三因素的相关系数如表 2 所示。

表 2 各因素之间的相关系数

性 状	单株产量 (Y)	单株穗数 (X_1)	穗粒数 (X_2)	千粒重 (X_3)
单株产量 (Y)	1	0.888 4**	0.564 1	-0.501 3
单株穗数 (X_1)		1	-0.639 4*	
穗粒数 (X_2)			1	-0.501 2
千粒重 (X_3)				1

注: $r_{0.05,8} = 0.632$, $r_{0.01,8} = 0.765$ 。

从表 2 不难看出,单株产量与单株穗数呈极显著正相关,单株穗数与千粒重呈显著负相关。这种相关关系是因素之间 10 个分量彼此关系的总和,具有一定程度的灰色性,反映不出其中所包含的多种相关趋势。因此,需要对其作进一步分析。

2.2 各因素观察值的同异反趋势划分

据表 1 数据和上述介绍的方法,可将单株产量、单株穗数、千粒重分别均分为小、中、大 3 个子区间:

单株产量: 小: [14.4, 16.767]; 中: (16.767, 19.134); 大: (19.134, 21.5)

单株穗数: 小: [8.4, 10.133]; 中: (10.133, 11.866); 大: (11.866, 13.6)

千粒重: 小: [34.1, 36.57]; 中: (36.57, 39.04); 大: (39.04, 41.5)

于是,由表 1 和上述区间值,得到各因素同异反趋势如表 3 和表 4 所示。

表 3 单株产量 Y 随单株穗数 X_1 变化的同异反趋势

品 种	1	2	3	4	5	6	7	8	9	10
单株产量 Y	大	中	大	小	中	小	小	中	小	小
单株穗数 X_1	大	大	大	中	大	中	中	大	小	中
同异反判断	同	异	同	异	异	异	异	异	同	异

表4 单株穗数 X_1 随千粒重 X_3 变化的同异反趋势

品种	1	2	3	4	5	6	7	8	9	10
单株穗数 X_1	大	大	大	中	大	中	中	大	小	中
千粒重 X_3	小	大	中	大	小	中	中	中	大	中
同异反判断	反	同	异	异	反	同	同	异	反	同

由表3和表4统计同异反子集数, 得单株产量 Y 随单株穗数 X_1 变化的同异反联系度表达式和单株穗数 X_1 随千粒重 X_3 变化的同异反联系度表达式分别为:

$$\mu(X_1, Y) = 3/10 + (7/10)i + 0j = 0.3 + 0.7i \quad (1^*)$$

$$\mu(X_3, X_1) = 4/10 + (3/10)i + (3/10)j = 0.4 + 0.3i + 0.3j \quad (2^*)$$

式(1*)表明, 在参试的10个品种中, 有3个品种的单株穗数与单株产量呈正相关趋势是确定的, 而有7个品种的单株穗数与单株产量的变化关系是正相关趋势还是负相关趋势则是不确定的。此时, 相关系数 $r = 0.8884^{**}$ 可以分解为 $|ra| = |0.8884 \times 0.3| = 0.26652$ 部分的正相关和 $|rb| = |0.8884 \times 0.7| = 0.62188$ 部分的不确定相关。在此意义下, 可以完全确定的相关系数部分为0.26652, 其余 $0.8884 - 0.26652 = 0.62188$, 则属于灰色不确定的部分。

式(2*)则表明, 在参试的10个品种中, 有3个品种的单株穗数与千粒重呈负相关趋势是可以肯定的, 有4个品种的单株穗数与千粒重呈正相关趋势和3个品种的单株穗数与千粒重的不确定相关趋势则是不能肯定的。此时, 相关系数 $r = -0.6394^*$ 可以分解为 $|ra| = |-0.6394 \times 0.4| = 0.25576$ 部分的正相关、 $|rc| = |-0.6394 \times 0.3| = 0.19182$ 的负相关和 $|rb| = |-0.6394 \times 0.3| = 0.19182$ 部分的不确定相关。在此意义下, 可以完全确定的相关系数部分为 $-0.6394 \times 0.3 = -0.19182$, 其余的 $-0.6394 - (-0.19182) = -0.44758$ 则属于灰色不确定的部分。

应当指出的是, 联系度表达式中的同一度和对立度仅仅表示的是呈正相关和负相关的一种趋势, 而不是呈正相关和负相关的程度, 这是需要加以严格区别的。如在式(2*)中, 同一度为0.4, 对立度为0.3, 似乎总体上应该呈正相关, 其实不然。尽管在10个品种中有4个品种单株穗数与千粒重有呈正相关的趋势, 只有3个品种单株穗数与千粒重有呈负相关的趋势, 但因为这3个品种的负相关

程度可能要远远大于4个品种的正相关程度, 所以总体上仍会呈现出负相关性。

由此可见, 相关系数 r 中包含了可以确定的相关关系(正或负)以及不确定的灰色关系两个部分。它能否反映研究对象两者之间的真实关系, 需视其灰色程度而定。因此, 需要对其灰色度进行研究。

2.3 相关系数的灰色度分析

据式(4)和(5), 求得单株穗数 X_1 与单株产量 Y 之间相关系数的灰色度为

$$f_c(0.8884) = 1 - a = 1 - 0.3 = 0.7$$

单株穗数 X_1 与千粒重 X_3 之间相关系数的灰色度为

$$f_c(0.6394) = 1 - c = 1 - 0.3 = 0.7$$

这就是说, 单株穗数与单株产量之间相关系数的灰色度以及单株穗数与千粒重之间相关系数的灰色度均高达70%。表明这两个相关系数所揭示的性质当中只有30%是完全确定的, 可以充分肯定。而有70%则是不确定的, 需要附加一定条件或补充一定信息才能由灰变白, 转化为确定性。因此, 在应用中要慎重对待, 不能因为其已达到极显著或显著水平便确信不疑。而应当在测验其显著水平的同时, 考虑其灰色度, 只有既达到显著水平, 灰色度又低的情况下, 方可断下结语。

3 讨论

相关系数 r 是自变量与因变量中各自变分量与因变分量之间相关关系的总和。当各分量之间的相关趋势完全一致时, 相关系数反映的意义是比较客观的。但实际上, 由于各种各样的误差干扰, 各分量之间的相关趋势并不总是一致的, 这就使得相关系数在某种程度上具有一定的虚假性。从信息论的角度看, 这种虚假性应归因于信息不完全。因此, 我们把这种现象叫做相关系数的灰色性。

在此之前, 虽然尚无人明确指出过相关系数的这种灰色性, 但已有学者曾围绕相关系数的缺陷开展过一定研究。如余国祥^[6]就曾指出, 按传统数理统计理论求得的相关系数 r , 其补数 $1 - r$ 中仍存

在相关不相关的信息, 并利用集对分析方法^[5], 结合学生成绩分析进行了具体研究, 主张把相关系数 r 一般性地写成 $a + bi$ 的形式, 表示为 $r + (1 - r)i$, $i \in [-1, 1]$; 严爱兰^[7]等通过学生两门课程相关情况的研究后认为, 相关系数由“优优、中中、差差”3 部分测度合成。这些研究都在一定程度上指出了相关系数的不足, 也提出了很有借鉴意义的破题思路, 但均未给出其合理的解释。我们在此基础上, 从集对分析和灰色系统的角度对相关系数进行了重新审视和观察, 明确指出了相关系数的灰色性, 并运用同异反分析的方法分析了相关系数中灰色性产生的根源, 用同异反联系度解析了相关系数的构成, 从相关系数中分离出灰色部分, 并提出了同异反分析意义下的相关系数灰色度概念, 由此揭示出数理统计、灰色系统、集对分析三者之间的一种内在联系。在这里, 同异反分析起到了一种桥梁和纽带的作用。

结合本文实例, 我们可以看出, 在研究因素之间的相关关系时, 除应考虑相关系数的显著性测验外, 还应考虑相关系数的灰色度问题。简言之, 即只有在 r 既达到显著水平, 又满足较低灰色度 (姑且定为 0.3 以下) 的情况下, 得到的结论才是比较可靠的, 才能避免因 r 的误用而导致研究结果的欠妥。当然, 灰色度的容许值究竟定在什么范围更合适, 尚需作深入研究。相关系数的灰色性是一种不确定性。正如人们所希望的那样, 这种不确定性不是一成不变的, 它在一定条件下 (如不断补充信息) 可以向确定性的方向转化, 从而加强或削弱自变量与因变量之间的相关程度。在转化过程中, 不确定系数 i 起着承前启后或承上启下的关键作用。当 i 取正值时, 灰色不确定量向正相关的方向转

化; 当 i 取负值时, 灰色不确定量则向负相关的方向转化。至于 i 取何值, 则应视具体情况而定, 一般有顺势取值法、逆势取值法、计算取值法、随机取值法、特殊值取值法等^[5], 但这些取值法都是特定情况下的特殊处理法, 不具普遍意义。是否可以通过可拓学理论^[7]来加以解决, 有待进一步研究。

同异反灰色相关分析的关键是各自变分量和因变分量的同异反趋势划分, 划分恰当与否直接影响分析结果。这里采用均分为 3 确定各分量数值为“小”、“中”、“大”的方法, 虽然从联系度的表达式中能反映出两因素之间的相关趋势, 但并未反映出两因素之间的相关程度, 因此, 不能把“相关趋势”和“相关程度”混为一谈。正因为如此, 除了应用过程中要引起特别注意外, 还须作更深层的研究, 以期寻求更为合适的划分方法使之日臻完善。

参考文献:

- [1] 莫惠栋. 农业试验统计 [M]. 上海: 科学技术出版社, 1992.
- [2] 王福亭, 郭瑞林. 农业试验设计与统计分析 [M]. 北京: 农村读物出版社, 1993.
- [3] 郭瑞林. 作物灰色育种学 [M]. 北京: 中国农业科技出版社, 1995.
- [4] 邓聚龙. 灰色系统理论教程 [M]. 武汉: 华中科技大学出版社, 1990.
- [5] 赵克勤. 集对分析及其初步应用 [M]. 杭州: 浙江科学技术出版社, 2000.
- [6] 余国祥. 基于 SPA 的学生成绩相关研究 [J]. 浙江师范大学学报, 1997, 20 (3): 37-41.
- [7] 严爱兰, 张林凤. 基于 SPA 联系数的体育统计相关系数研究 [J]. 湖北体育科技, 1990, 18 (1): 59-63.
- [8] 蔡文, 杨春燕, 林伟初. 可拓工程方法 [M]. 北京: 科学出版社, 1997.

(上接第 169 页)

- [6] 李铁松, 张桥英. 嘉陵江的起源及流域自然环境研究 [J]. 四川师范学院学报 (自然科学版), 1999, (2): 105-108.
- [7] 严贤春, 苏智先. 水土保持与生态农业旅游技术研究 - 以四川省南充市嘉陵区木老乡生态农业旅游区为例 [J]. 生态学杂志, 2003, 22 (6): 157-161.
- [8] 肖笃宁. 国际景观生态学研究的最新进展 [J]. 生态学杂志, 1999, 18 (6): 75-76.
- [9] 张慧, 缪旭波, 孙勤芳. 景观生态学在农业景观生态规划中的应用 [J]. 农村生态环境, 2001, 17 (1): 29-32.
- [10] 宗跃光. 城市景观生态规划中的廊道效应研究 [J]. 生态学报, 1999, 19 (2): 145-150.
- [11] 唐亚, 谢嘉穗. 坡耕地的水土保持和持续耕作 [J]. 中国水土保持, 2002, (7): 29.
- [12] 徐峰. 观光农业景观设计 [J]. 林业建设, 2003, (2): 15-18.
- [13] 魏天兴, 朱金兆, 朱清科, 等. 黄土陡坡地农林复合经营设计与水土保持效益研究 [J]. 土壤侵蚀与水土保持学报, 1998, 4 (2): 82-87.
- [14] Forman R T T. Land mosaics: The ecology of landscape sand regions. Cambridge, U K: Cambridge University Press, 1995.
- [15] Shipley P R, J J Meisinger, A M Decker. Conserving residual corn fertilizer nitrogen with winter cover crops. Agron J, 1992, (84): 869-876.