

百分数和分级资料的统计分析

在日常科研工作中,很多资料需要用百分数表示。但如何分析百分数,许多人感到很混乱并因此时常造成许多误解。我们常会遇到这样一些问题:什么情况下对百分数实施反正弦转换?为什么要进行数据转换?怎样判断百分数的客观性?对目测的百分数进行方差分析有效吗?分析包括0%的资料是否有效等等,诸如此类。本文介绍并简答上述有关问题。

众所周知,方差分析是对试验数据进行统计分析的有效手段。但是,方差分析是建立在一些假定基础上的,其基本的假定可归纳为可加性、正态性、同质性和独立性四点。许多百分数资料的一个典型特点是,百分数的大小与其误差直接相关。即在样本容量一定时,两极端的百分数(接近0或100%)方差较小,而中间(50%左右)较大。这种情况直接违背了各处理方差的同质性假定。对百分数进行反正弦转换,又称角度转换,主要是使具有小方差的小百分数值增大,大百分数值减小(例如,5%的反正弦值是12.92,90%的反正弦值为71.56),因此,可改进具有不同百分数值处理的均方异质性,使转换后的数据满足方差分析的假定条件。

反正弦转换是通过求角度得到的:角的正弦是该百分数的平方根,计算公式为 $Q = \sin \sqrt{p}$ 或 $Q = \arcsin \sqrt{p}$ 。进行反正弦转换的具体方法有两种,一是直接查百分数反正弦转换表;一是根据公式,用计算器计算出变量值。

是否需要反正弦转换的一般原则是,如果观测数据的大部分值在30—70%之间,不必转换,可直接进行方差分析。如果大部分观察值超出30—70%的范围,即小于30%或大于70%,则需对全试验数据反正弦转换,然后对转换后的数据进行方差分析。

在方差分析过程中,一切显著性测验、平均数的计算和平均数的多重比较,都应根据转换后的数据进行,而不是根据原始数据。表达试验结果时需将转换后数据的平均数作“反转变换”为原来的尺度。对进行了反正弦转换的资料应予说明。兹举一实例计算如下:

有一个A、B、C、D四个大白菜品种的品比试验,随机区组设计,五次重复。统计其病毒病发病率,得结果于表1,试分析品种间抗病性的差异。

表1中调查的原始数据的变异幅度为0.0~84.6%,大部分数据小于30%,所以该资料宜作反正弦转换。转换结果列于表1右边五列,对此数据方差分析的结果见表2。F测验显著,表明这四个品种的抗病性间存在显著差异;LSR法多重比较(表1)。并将各品种反正弦值的平均数反转为百分率(表1)。本分析结果指出,A品种平均发病率为1.8%,B为2.5%,显著低于C和D两品种;表明A、B两品种对病毒病的抗性显著高于C和D。

表1 四个白菜品种病毒病发病率

品种	调查的百分数					$\sin \sqrt{p}$					平均数	反转换为%
	1	2	3	4	5	1	2	3	4	5		
A	0.8	3.8	0.0	6.0	1.7	5.13	1.24	0.00	4.18	7.49	7.61	1.8
B	4.0	1.9	0.7	3.5	3.2	1.54	7.92	4.80	0.78	0.31	9.07	2.5
C	9.8	56.2	66.0	10.3	9.2	8.24	8.56	4.33	8.72	7.64	31.50	27.3
D	6.0	79.8	7.0	84.6	2.8	4.15	3.29	5.36	6.89	6.3	33.87	31.1

表2 表1转换后资料的方差分析

变异来源	自由度	平方和	均方	F
区组	4	1430.08	357.52	3.57*
品种	3	2982.71	994.24	
误差	12	3337.90	278.16	
总	19	7750.69		

通常,我们遇到的百分数可分为三类。第一类,也是我们最常处理的百分数,即表示一些成分百分含量的数据,如糖的含量、可溶性固形物含量、干物质含量、蛋白

质含量等等。这类数据的数值范围有限,一般不会出现极值,而且这类变数通常趋于正态分布。对这类测定值一般不必转换。

第二类百分数即成数,是指结实率、发芽率、杀虫率、病株率等这类统计数。这类数据符合二项分布,具有两端平均数(接近0或100%)的方差较小,而中间(50%左右)的方差较大的性质。因此,如果观察值的范围符合前述的正弦转换原则,就要对此进行转换。

如果观察值的范围在0~20%或者80~100%之间,亦可应用平方根转换。当百分数都小于20时,可直接转换为平方根,公式为 $y' = \sqrt{y}$ (如果包括0,则用 $y' = \sqrt{y+1/2}$);当百分数都大于80时,需先从100减去之,再作平方根转换。

第三类是指那些根据分级而得到的百分数。分级资料在大量的文献报道中随处可见,作者一般是根据自己的情况采用相应的分级制,没有统一的分级体系。常用的一些分级可以量化。例如病害级别的划分一般是根据植株受病菌感染部分的百分率确定的,级别中的每一级代表一个百分数的范围。如茄子黄萎病的分级:0级表示无病或基本无病,1级为病叶面积 $\geq 25\%$,2级病叶在26—50%之间,3级病叶在51—75%之间,4级病叶面积 $> 75\%$ 。当然,分级也不一定必须包括0—100%的整个范围。如落叶的级别分1—5级,1级表示25%落叶,5级则没有落叶。

上述这类分级,一般级差都是等百分数增加。以此统计分析时,若观察值超出30—70%的范围,就有必要进行反正弦转换。因为一般地,这类分级数据处理内两端平均数的方差小于中部平均数的方差。这实际上是显而易见的:因为人眼对近极端处较高或者较低百分数的分辨力高于中间部位。例如,目测除草率时,区分95%和

99%并不困难,但却很难区分开45%和49%。

为了使肉眼能等程度地区分开每一级,horsfall等提出了测定植株病害目测分级的另一种方法:当病组织 $< 50\%$ 时,观测病部;病害严重度 $> 50\%$ 时,观测健部组织。其原理是,目测的敏锐度与被观测物体的对数成比例。这种分级是从中间开始的,每一级大约是前一级的一半。向下级别是,50—25%、25—12%、12—6%、6—3%、3—0%,向上的级别是50—75%、75—88%、88—94%、94—97%、97—100%。这种分级法使两极端的数据比中间的数据得到了较大的加权,在一定程度上弥补了典型百分资料的方差异质性。因此,统计分析时不必转换数据。

由分级得到百分数一般都是目测的结果,其是否客观?有人认为,根据人的判断或感官评定的结果是主观的,实验室的仪器比人的感官更准确。但实际并非如此,只有受人的偏见影响的观测才是主观的,而人的偏见可以避免并可进行测定;对食品等风味的鉴别,人的感官一般比实验仪器准确。因此,只要不存偏见,目测的百分数同样是客观的。在分析之前,检查数据是否违背了方差分析的假定条件,如果必要,进行数据转换,然后即可进行有效的方差分析。

除了对数和倒数,分析包括0%的资料基本上不存在错误。

总之,分析处理分级资料也并不复杂。首先,建立一套行之有效、观测者容易分辨的分级制,这种分级的级差可以是不等量的。第二,应用分级标准调查时,努力避免偏见。第三,检查数据的方差是否同质及平均数与方差是否独立;如果不满足方差分析的假定条件,采用适当的变量转换法变换数据,然后则可进行方差分析。(黑龙江省农科院园艺研究所 哈尔滨 150069)

黑龙江省农科院园艺研究所白菜研究室主任李光池副研究员 向广大新老用户致意

在各用户的大力支持下,我室的大白菜和萝卜育种研究,进展迅速,谨致谢意,并真诚希望合作永恒。

我室选育的大白菜和萝卜新品种,以产量高、抗病性强、品质优、适应性广、纯度好深得广大用户信赖,品种有龙白一号、黄籽、九号,尤其是最新推出的大白菜龙辐二牛心和龙协白三号、87—2,深受菜农喜爱。我室还是省种子公司的萝卜原种定点生产单位。

我室竭诚为广大用户提供大白菜、萝卜的优良品种,采用合同、电报、信函、面议等形式订货,汇款、现金

等方式结算,大量购种、联合经营、联合繁种、原种转让等合作方式均可商议。欢迎全国新老用户来人来函洽谈业务,欢迎国内外各界朋友通力合作。

本室地址:哈尔滨市哈平路义发源(在火车站乘338路汽车到义发源下车回走200米即到)。电话:63603、63606、62369 转白菜室或李光池、鹿英杰家。电挂:4951,邮编:150069。开户行:哈市农行跃进营业所。帐号:431454004。联系人:李光池、鹿英杰。