

利用自助法进行种子室内检验的初步研究

杨素青 刘晓明 张力军

(齐齐哈尔市种子公司)

李 青

(齐齐哈尔林业学校)

种子检验是一项政策性、技术性很强的工作,是种子工作和农业生产中必不可少的一个环节。影响种子检验数据准确性的因素主要有抽样、试验、认识三大误差,而抽样误差位于三大误差之首。抽样理论告诉我们,抽取样品的数量越多,结果越准确,但这样不仅加大了种子检验的工作量,还浪费了一定的种子。针对上述问题,本文试图利用自助法进行这方面的探讨,以求在保证检验精度的条件下减少抽取样品的数量,达到事半功倍的目的。

1 材料来源

材料来自 1987~1990 年黑龙江省三个地区的室内检验数据(见表 1)。对表 1 中各品种每个总体单元(即每袋种子)的净度、水分、芽率三项检验指标按检验规程进行了测定,得到了每个总体单元的水分、净度、芽率的实际检验数据(以百分数表示),并分别计算了三项检验指标的均值。限于篇幅,这些数据从略。利用这些数据按自助法原理进行分析研究。

表 1 各品种统计表

年 度	1987	1988	1988	1989	1990
试验地区	齐齐哈尔	嫩江	齐齐哈尔	龙江县	齐齐哈尔
作 物	玉 米	玉 米	高 粱	玉 米	玉 米
品 种	合玉 14	7024	齐 14A×齐康千	东 46	龙单一
总体单元数(袋数)	61	62	57	43	104

2 自助法的基本原理

自助法(Bootstrap)是在 1979 年由 Efron 提出来的,是一种非参数统计方法。它的基本原理是:设从某一总体中抽取 M 个观测数据,组成一个样本,称为初始样本。使用随机数发生器,等概、独立、有放回地从 M 个观测数据所组成的集合中抽取 M 个点(即 M 个数据),组成一个新的点集,称为一个自助样本。在自助样本中,初始点集中的点,有的被抽中 0 次,有的被抽中 1 次、2 次等等。即自助样本是初始样本的子集。照此重复抽取 N 个自助样本,对每个自助样本求出一个我们所关心的统计量($\hat{\theta}_i$)。本文所要讨论的统计量($\hat{\theta}_i$)是各检验指标(净度、水分、芽率)的均值。则检验指标均值自助估计量($\hat{\theta}$):

$$(\bar{o})=\frac{1}{N}\sum(\hat{o}_i) \dots\dots\dots (1)$$

将 N 个自助样本统计量由小到大排序,即 $\hat{o}_1\leq\hat{o}_2\leq\dots\leq\hat{o}_N$, 设 a 为这一序列的第 16 个百分点数(或第 1 585 个万分位数),b 为这一序列的第 84 个百分点数(或第 8 415 个万分位数),则定义

$$\sigma^{(n)}=\frac{b-a}{2} \dots\dots\dots (2)$$

为统计量 (\bar{o}) 的标准差自助估计量。

$$p_{估}=1-\frac{t_{\alpha}\cdot\sigma^{(n)}}{(\bar{o})}\times100\% \dots\dots\dots (3)$$

为统计量 \hat{o} 的精度自助估计量(本文 t_{α} 是以 95% 的可靠性查表而得)。

3 计算过程及结果分析

3.1 计算过程

根据自助法原理可知,当 M 和 N 确定时,检验指标均值自助估计量 (\hat{o}) 和自助估计精度 $(p_{估})$ 也是确定的;当 M 或 N 变化时, (\hat{o}) 和 $p_{估}$ 也随着变化。为便于弄清 (\hat{o}) 和 $p_{估}$ 随 M 或 N 的变化规律,本研究中,M 的取值分别为 3,4,5,6,7,8,9,10,12,14,16,18,20,25,30;N 的取值分别为 20,40,60,80,100,150,200 计算了 \bar{o} 和 $p_{估}$ 。研究中发现,利用自助法估计各检验指标的均值 (\hat{o}) ,估计精度 $(p_{估})$ 随 M 或 N 的变化有一定的规律性,并且还发现,不同总体的估计精度随 M 的变化规律一样,即 $p_{估}$ 随 M 的增加而增大,当 M 达到一定数值时, $p_{估}$ 趋于稳定。

表 2 玉米 7024 水分含量的精度比较

精度 N		M	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
20	$p_{实}$		77.34	70.34	85.34	92.31	96.46	99.98	99.84	98.41	99.38	99.85	99.71	99.83	99.38	99.87	99.47
	$p_{估}$		76.07	68.23	82.41	90.56	94.44	99.12	98.87	97.47	98.82	98.95	98.56	99.01	98.75	99.03	98.95
40	$p_{实}$		79.80	75.82	84.25	89.25	95.31	99.76	99.73	98.81	99.33	99.88	99.85	99.71	99.37	99.61	99.58
	$p_{估}$		74.28	74.25	83.11	86.01	92.62	97.93	98.59	97.80	98.13	98.73	98.52	99.00	98.74	99.15	98.57
60	$p_{实}$		78.34	80.63	79.21	91.56	95.45	99.78	99.81	99.09	99.36	99.83	99.55	99.83	99.57	99.55	99.35
	$p_{估}$		80.63	78.26	78.46	89.34	91.35	98.65	98.23	98.20	97.81	98.68	98.36	98.79	98.75	98.87	98.97
80	$p_{实}$		82.35	81.30	86.04	92.38	97.51	99.88	99.91	98.82	99.22	99.92	99.60	99.90	99.36	99.60	99.53
	$p_{估}$		79.40	78.35	83.25	91.47	95.32	97.85	98.51	97.61	98.50	98.54	98.36	98.76	98.84	98.85	98.93
100	$p_{实}$		82.43	80.60	81.12	91.38	96.34	99.82	99.78	98.75	99.27	99.85	99.62	99.93	99.49	99.81	99.60
	$p_{估}$		68.24	79.45	79.35	89.45	90.41	98.09	98.66	87.86	98.55	98.31	98.76	98.62	98.85	98.75	98.83
150	$p_{实}$		79.65	81.83	85.64	86.45	93.56	99.73	97.36	98.85	99.19	99.91	99.74	99.85	99.48	99.65	99.52
	$p_{估}$		75.23	78.64	83.25	88.02	91.30	98.25	96.56	97.87	98.18	98.49	98.60	98.76	98.85	99.00	98.84
200	$p_{实}$		77.41	80.36	80.31	92.30	92.89	99.73	99.83	98.87	99.20	99.94	99.70	99.87	99.46	99.66	99.58
	$p_{估}$		72.02	76.38	78.25	89.21	91.78	98.33	98.23	97.74	98.34	98.40	98.52	98.69	98.72	98.90	88.97

现将某一总体,如 7024 玉米种子水分的估计精度 $(p_{估})$ 随 M 或 N 变化情况列于表 2,即可看出每个总体的估计精度 $(p_{估})$ 随 M 或 N 变化的规律。

同时,我们把各总体检验指标的算术平均值 (\bar{o}) 作为真实值,按(4)式计算了实际精度 $(p_{实})$ 。

$$p_{\pi} = 1 - \frac{|\bar{0} - \hat{0}|}{\hat{0}} \times 100\% \dots\dots\dots (4)$$

式中: $\bar{0}$ ——各检验指标算术平均值

$\hat{0}$ ——按自助法计算的均值

不同总体的实际精度(p_{π})随 M 或 N 的变化规律一样,也是 p_{π} 随 M 的增加而增大并逐渐趋于稳定,从表 2 中即可看出。

3.2 结果分析

3.2.1 从表 2 中看出,自助法估计各检验因子的估计精度和实际精度是随初始样本单元数(M)的增加而逐渐增大,直到 $M \geq 7$ 时,精度变化趋于稳定。

3.2.2 当 $M \geq 7$ 时,除东 46 品种(表中未列)的芽率估计精度为 84.24%和实际精度 86.08% ($M=7$ 时)外,其余的估计精度和实际精度均大于 90%。

3.2.3 自助样本单元数(N)的大小对精度高低的作用不显著。据笔者研究所知,当 $N < 60$ 时,自助样本单元不呈正态分布,只有当 $N \geq 60$ 时,此样本单元才呈正态分布,并且标准差也趋于稳定。所以应用自助法估计各检验因子均值必须在 $N \geq 60$ 的条件下进行。

4 小结

从研究过程和结果看出,利用自助法估计种子的净度、水分、芽率,在初始样本单元数为 7 和自助样本单元数为 60 的条件下,就可以得出估计精度和实际精度均大于 90%的满意结果。它在种子室内检验工作中确是一种理想的方法,值得在生产中推广和应用。

参 考 文 献

- 1 黄承明译. Bradley Efron 计算机与统计理论. 应用数学与计算数学, 1980(5)
- 2 王素玲等. 自助法在森林抽样调查中的应用. 辽宁林业规划设计, 1988(3)
- 3 郎奎健等. IBMPC 系列程序集. 中国林业出版社, 1989

启 事

本刊编辑部尚有 1993 年、1994 年《黑龙江农业科学》及北方春麦区小麦论文专辑(2.00 元/册)、大豆重迎茬问题研究专辑(1.00 元/册)和农化产品应用技术论文专辑(1.30 元/册)。有需要者,请将款汇至本刊编辑部,款到寄书;另有漏订 1995 年《黑龙江农业科学》的读者,请将款汇到编辑部直接订阅。

黑龙江农业科学编辑部