

病虫害防除所の現場で想定される データサンプリングと解析手法

独立行政法人 農業・食品産業技術総合研究機構 中央農業総合研究センター
情報利用研究領域 光永 貴之

1. はじめに

現在、発生予察業務は病虫害防除所における最も中核的業務の一つである。発生予察は、ある病虫害の近い将来における発生を現在の発生状況から精度よく予測することにより、その病虫害の効率的防除の指針とすることが目的であるので、可能な限り正確であることが望ましい。そのために最も単純な手段は、できるだけ多くの地点で継続して調査を行うことであるが、実際には、人員や資材等の制約の中で大規模調査は難しい。つまり、いかにして少ない観察結果から、現在、そして将来の発生状況を正確に予測できるかが重要になってくる。

本稿では、発生予察業務におけるデータサンプリングの考え方及びこれに対応した解析手法、現場での統計的判断について簡単に紹介する。実際の予察業務の一助になれば幸いである。

2. データサンプリングの注意点

調査地点数の決定は予察精度を左右する最も重要な問題であるが、対象となる病虫害の性質によって異なった考え方をしなければならない。対象の病虫害が毎年必ず発生し、発生の有無ではなく、発生量の予測が重要である場合には、比較的少数の調査地点で問題がない。このような病虫害では調査データの平均値の変動が重要になるのだが、平均値は再現性が高いためである。このような病虫害はほとんどの場合、古くから知られたメジャーなものであり、予察手法も慣行のもので問題は少ないであろう。

一方、発生そのものが重要なもの、あるいは発生量が少なくても農作物に甚大な被害を与える病虫害の予察は、多大な労力を必要とすることが多い。例えば、激しい症状を発現する病害や大型カメムシ類などは、地域内での分布はランダム分布（ポアソン分布）に従うことが多く、これを元に調査地点数を決定することができる。例としてある害虫の現時点での地域での個体数密度が平均1頭/圃場であったと仮定する。この時に一か所の調査でその発生を確認できない確率はポアソン分布の0項から推定される。

$$P(n=0, \lambda=1) = \frac{1^0}{0!} e^{-1} = \frac{1}{e} \approx \frac{1}{2.718} \approx 0.369$$

したがって予察業務により、この害虫の発生を95%以上の確率で検出するには $0.369^n < 0.05$ を満たす n を見つければそれが必要調査地点数となる。今回の n は $3.00487\cdots$ となるので、最小調査地点数は4となる。この考え方では、調査時点での平均発生数が小さければ小さいほど必要調査地点数は増大する。例えば、平均0.1頭(株)/圃場では30地点、平均0.01頭(株)/圃場では300地点が必要となる。上記の考え方を基にすれば対象としている病虫害の(現時点での)許容発生密度を決定することにより、必要調査地点数を計算することができるが、労力等の問題から何らかの妥協(例えば、検出確率を90%に下げる)が必要になる可能性もある。

3. データサンプリングに対応した統計手法と統計的判断

予察データは通常、平年値との違いを統計的判断によって解釈する。したがって、収集されたデータに応じた統計手法を知っておくことが重要である。発生量が重要となる病虫害(調査ごとにおおよそ数十個体(株)を確認)についてはパラメトリック検定(t検定や分散分析等)によって例年の発生量との比較が可能である。一方、発生量が極端に少ない病虫害については、ポアソン回帰、あるいはロジスティック回帰等の正規分布を仮定しない検定手法を用いる方が望ましい。

しかしながら、結果の解釈については注意を要する。統計的仮説検定は“A群とB群が違う”ことを示すことには適しているが、有意差がなかったからと言って“同じ”ということを担保するものではない。すなわち平年値との比較で有意差があれば、“今年は発生量が多い(少ない)”という判断ができるが、有意差がないからと言って“今年の発生量は平年並みである”とは言えない。このように有意差だけでは判断できない場合には、平年値のばらつきの範囲内に本年値のばらつきが収まっている等の付加的情報を用いる必要がある。