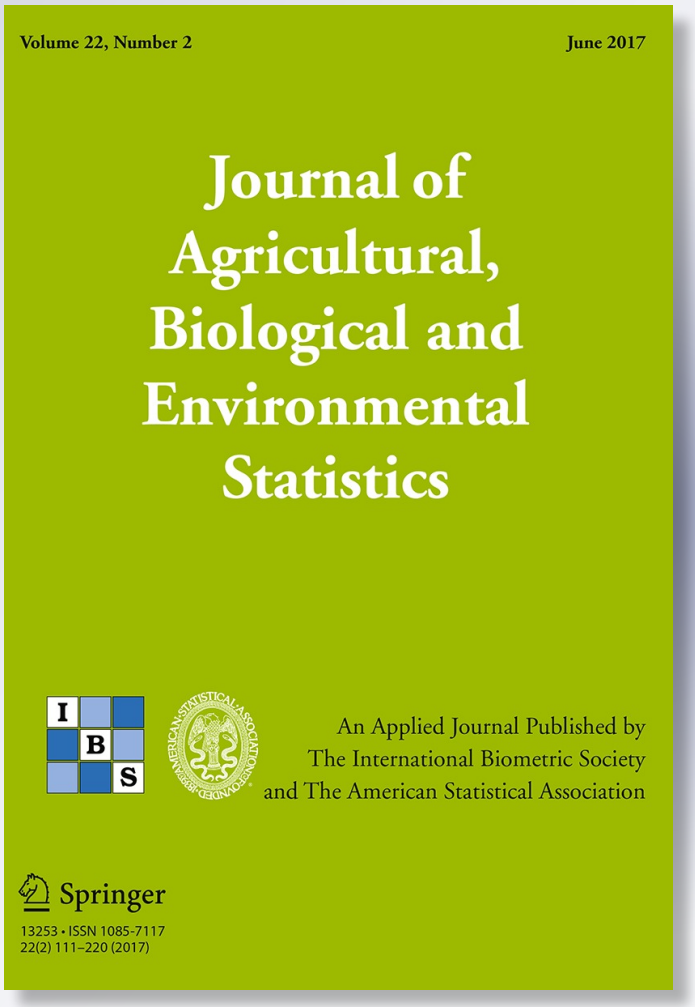
A Statistical Method to Construct Confidence Sets on Carrion Insect Age from Development Stage

Lynn R. LaMotte, Amanda L. Roe, Jeffrey D. Wells & Leon G. Higley

Journal of Agricultural, Biological and Environmental Statistics

ISSN 1085-7117
Volume 22
Number 2

JABES (2017) 22:161-171
DOI 10.1007/s13253-017-0280-y



Springer

A Statistical Method to Construct Confidence Sets on Carrion Insect Age from Development Stage

Lynn R. LaMotte, Amanda L. Roe, Jeffrey D. Wells, and Leon G. Higley

The age of a carrion insect associated with a corpse may represent a minimum postmortem interval. No method has been proposed before for constructing a confidence set on age based on development stage modeled as a categorical response. This paper illustrates the application of exact p values, first developed for succession data, to construct a confidence set on a carrion insect's age based only on its development stage. It uses published development data for *Lucilia sericata*, with individuals reared at different temperatures pooled into sets of similar age as indexed in accumulated degree hours. Rates of coverage of true ages, assessed using each insect as a singleton holdout sample, were greater than the nominal 95% level.

Key Words: Calibration; Inverse prediction; Categorical response; Outlier detection; Forensic entomology; *Lucilia sericata*.

1. INTRODUCTION

A common forensic entomological analysis involves estimating the age of a carrion insect associated with a corpse (Catts 1990). If circumstances suggest, as is usually the case, that the individual insect specimen was deposited by its mother on the victim following death, then the implication is that the victim was dead for a time period at least equal to, but perhaps greater than, the age of the insect specimen. Catts (1990) referred to this concept as a minimal postmortem interval, now more commonly referred to as a minimum postmortem interval (PMI_{min}) (see Goff 1993; Amendt et al. 2007; Villet et al. 2010).

An analyst should be able to objectively express the uncertainty concerning any forensic science conclusion (National Research Council 2009). When statistical methods are used,

Lynn R. LaMotte (✉) LSU Health Sciences Center, New Orleans, LA, USA
(E-mail: *llamot@lsuhsc.edu*). Amanda L. Roe College of Saint Mary, Omaha, NE, USA
(E-mail: *aroe@csm.edu*). Jeffrey D. Wells Florida International University, Miami, FL, USA
(E-mail: *jedwell@fiu.edu*). Leon G. Higley University of Nebraska - Lincoln, Lincoln, NE, USA
(E-mail: *lhigley1@unl.edu*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 22, Number 2, Pages 161–171
DOI: 10.1007/s13253-017-0280-y

this takes the form of levels of significance or levels of confidence. Methods have been proposed for calculating confidence limits for an insect specimen's age based on a continuous quantitative response(s) such as body length or weight (see Wells and LaMotte 1995; Ieno et al. 2010; Baque et al. 2015; LaMotte and Wells 2015, 2016). However, size can vary considerably between individuals of the same age (Wells and LaMotte 2001), and body length is influenced by specimen preservation method (Tantawi and Greenberg 1993; Adams and Hall 2003; Midgley and Villet 2009). For such reasons, a forensic entomologist may need to estimate age based only on development stage, which is categorical (see Dadour et al. 2001; Gaudry et al. 2001, and Huntington and Higley 2008).

No probability-based statistical methodology has been proposed for inferring carrion insect age from a categorical response such as development stage. (It has been used as a predictor, but not as a response (Ieno et al. 2010).) LaMotte and Wells (2000) devised an exact method for a similar setting, inferring the time interval a corpse was exposed to carrion insects based on species occurrence data. In this paper, we will show how the exact probability basis derived there can be applied to construct confidence sets on PMI_{min} based on categories of insect development.

The data for this illustration are particularly important to forensic entomology. They come from the large development study by Roe and Higley (2015) on *Lucilia sericata*, a species of carrion-feeding fly that is important in forensic applications. We assess the statistical performance of this approach in terms of the confidence sets' coverage rates, both of the true value and of false values, using each subject in the data set as a singleton holdout sample.

2. MATERIALS AND METHODS

2.1. Development Data

We used the *Lucilia sericata* development data recently published by Roe and Higley (2015). In that study, for most of the temperature–age combinations examined, four rearing cups each with approximately 20 eggs of equal age (cohorts) were assembled. More than four cohorts were observed for some of the oldest sampling ages, and the surplus ones were used to observe adult emergence. Because of pre-sampling mortality, sample sizes tended to decrease with age. There were 1024 cups from the 239 age-temperature combinations. The complete data set includes development stages of 11703 individual *Lucilia sericata* specimens at 239 distinct combinations of eleven incubator temperatures and 197 ages (rounded to hours since oviposition).

2.2. Pooling Observations by Age in Accumulated Degree Hours (ADH)

The effect of temperature on blow fly development rate with age is profound. The combined effects of age and temperature on development may be modeled in terms of a single index, accumulated degree hours (ADH) (Catts 1990). Calculating ADH requires a temperature threshold, and we selected 10°C because it is within the range of values estimated for

L. sericata (Roe and Higley 2015; Wall et al. 1992). For the analyses reported here, we combined data by groups of specimens within brackets of ADH. We also excluded observations at the highest and lowest rearing temperatures (7.5 and 32.5°C).

Although some of the 201 distinct ADHs in the data set corresponded to multiple temperatures, many had numbers (n) of insect specimens, combined over cups and temperatures, that were too low for this kind of analysis. For example, with four categories of development stages, it is not possible to detect a difference at the 5% level of significance unless $n \geq 22$; for six categories, unless $n \geq 37$ (LaMotte and Wells 2000). To avoid so many with small sample sizes, we combined ADHs into brackets (disjoint ranges of values) such that the greatest in a bracket was not more than 6% greater than the least (see Table 3).

2.3. Definition of Categories of Development Stages

Roe and Higley (2015) designated development stages into fifteen categories, from egg to adult. We considered two less refined categorizations, to four and six stages. For presentation here, we excluded postfeeding stages and combined the younger stages into four categories. They are designated here and defined as follows: E (egg and pharate first larval instar); L1 (egg hatching, first larval instar, and pharate second larval instar); L2 (first larval molt, second larval instar, and pharate third larval instar); and L3-f (second larval molt and feeding third larval instar). For casework, the selection of life stages to be covered by the model would be up to the investigator. The choice of number of development categories or laboratory temperatures to include does not affect the logic of the methodology illustrated here. We refer to the assignment of these four stages to specimens as DS (for "developmental stage") and to the four categories collectively as DSs.

Roe and Higley's (2015) purpose was to examine the relationship between development rate and temperature, not to support the application we present here. In our concluding remarks, we discuss aspects of experimental design we think would be helpful for those who aim to produce an age-prediction model. However, the data provide a good example for illustrating this statistical method. They are among the most comprehensive yet produced, and particularly important for statistical inference, the authors employed an unbiased sampling scheme (Wells et al. 2015).

2.4. Statistical Methodology

The objective of constructing a confidence set on age from DS is addressed by computing a p value comparing the observed mystery specimen ("MS") DS to the distributions of DSs for the training data ("TD," i.e., specimens for which both DS and age are known) in each of the ADH categories.

This rationale is the same in conventional statistical methods for *inverse prediction* or *calibration*. There, training data are used to fit a model for the quantitative response y in terms of a quantitative condition x. In its conventional form, the model is linear, e.g., $y = \beta_0 + x\beta_1 + \epsilon$. Then, given a known response y_* for which the value x_* of the condition is unknown, the objective is to infer what values of x could reasonably have yielded y_*. To do so, y_* is tested as an outlier at the α level of significance ($\alpha = 0.05$

here) for each potential value of x_*. Those values of x_* for which the test does not indicate that y_* is an outlier comprise a $100(1-\alpha)\%$ confidence set on the true x_*. (This follows from Theorem 4 (i) in Lehmann 1986, p. 90.) A value x_* is in this set if and only if y_* is contained in the *prediction interval* on y at x_*, hence the name of the method.

In the setting of simple linear regression, under the conventional assumptions of independence, homoscedasticity, and normality, the statistical properties of inverse prediction are straightforward to establish. In particular, the confidence set on x_* is an interval, and the probability that it covers the true value is the stated level of confidence. Achievement of the exact coverage probability is guaranteed by exact p values for the outlier tests; each potential x_* is in the confidence set if and only if the corresponding outlier test p value is ≥ 0.05.

In principle, the same rationale can be used when the response is categorical. The crucial step is to test the observed MS category of the response as an outlier from the probability distribution of categories at a given x_* as estimated with the training data. Exactness of the p value of that test guarantees the coverage probability of the resulting confidence set on the unknown condition x_*.

The probabilistic foundation for this exact p value is described in LaMotte and Wells (2000). Here, the condition corresponding to x_* is ADH. ADHs for which this p value is less than 0.05 are rejected as being untenable for the given specimen. The set of tenable ADHs (those for which the p value is not less than 0.05) constitutes a 95% confidence set on ADH of the MS. Although 0.05 is the conventional default level of significance, tables in LaMotte and Wells (2000) enable an investigator to choose other levels.

While we can compare a mystery specimen to the frequency distribution observed at an ADH represented in the training data, we do not attempt here to impute what the conclusion would be at ADHs not represented in the training data. Clearly, modeling how the distribution of stages changes with ADH can be important and useful. But no model is the true model, and formulating, fitting, and assessing models adds layers of arbitrary complexity that would confound the objective here, which is to illustrate, and assess the performance of, these exact p values.

The comparison between TD and the MS is in the form of a separate contingency table for each ADH (or bracket of ADHs) in the training data. A numerical example will serve to illustrate how a p value is found in this setting. One of the ADH combinations resulted from cohorts sampled after 79 h in the incubator at 22.5°C, plus 2 h at room temperature (20°C), which together translate to ADH$= 2(20-10)+79(22.5-10) = 1007.5$ degree hours. Of 35 specimens from three cohorts (cups), 20 were in stage L2, 15 were in L3, and none was in E or L1. This frequency distribution, (0, 0, 20, 15) in the four categories, constitutes the training data at this ADH. Consider a MS that is in the L1 stage. To address the question, whether it comes from the same probability distribution as the training data (TD) at 1007.5 degree hours, its distribution (0, 1, 0, 0) is compared to the training data distribution in a 4×2 contingency table (Table 1).

Conventional methods for computing a p value for comparing the MS frequency distribution to the TD distribution, the Chi-squared test and Fisher's exact test, are inaccurate

Table 1. Frequency distributions of 35 training data (TD) specimens and one mystery specimen (MS) over four development stages (DSs).

DS	TD	MS
E	0	0
L1	0	1
L2	20	0
L3-f	15	0

The exact p value comparing the MS distribution to the TD distribution is 0.0311.

Table 2. Frequencies for the four DSs (column TD) are those of the 195–200 ADH bracket in the TD shown in Table 3.

DS	TD	MS_1	MS_2	MS_3	MS_4
E	345	1	0	0	0
L1	34	0	1	0	0
L2	0	0	0	1	0
L3-f	0	0	0	0	1
p value		>0.05	>0.05	<0.05	<0.05

Columns MS_1–MS_4 show frequency distributions of single potential MSs from each of the four DSs.

The p value at the *bottom of each column* was computed as described in LaMotte and Wells (2000) or using Table 4 with $n = 379$. With $c = 4$ DSs, a MS in any DS with TD frequency ≤ 9 gives a p value <0.05 and leads to rejection of the MS coming from ADH 195–200.

because the MS sample size is 1. An accurate p value can be computed under the null hypothesis, that the $35+1$ observations are sampled independently from the same 4-category Bernoulli distribution. LaMotte and Wells (2000) give its mathematical derivation. It is an exact probability, not an approximation, and so it is not limited by sample size or empty cells. Although it entails multi-dimensional constrained optimization, they showed that its value can be computed to within 10^{-4} by a simple formula that requires only a one-dimensional optimization. Although there is not a closed-form expression for this probability, the algorithm can be programmed readily. We computed the p values shown in this paper in that way.

LaMotte and Wells (2000) provided a table (their Table 3, extended here in Table 4) to be used in lieu of the optimization, showing the least TD sample size n such that the training data frequency of the MS category ($x = 0$ in Table 1) would result in a p value less than 0.05. From that table, that minimum sample size is 22, and the training data sample size here is 35, and so we may conclude that the p value is less than 0.05 without actually computing the p value.

To further illustrate the statistical test, note that each row in Table 3 comprises the training data column in a contingency table like the one shown in Table 1. As described in the caption of Table 3, a frequency framed or unframed in Table 3 indicates a mystery specimen DS that would be rejected or not rejected, respectively, as coming from that ADH or bracket of ADHs. Table 2 illustrates the four comparisons (one for each DS) for the 195–200 ADH bracket in Table 3.

Table 3. Frequency distributions of four DSs among training data specimens within each ADH bracket.

ADH	DS				*n*
Bracket	E	L1	L2	L3-f	
100–105	154	[0]	[0]	[0]	154
107–110	402	[0]	[0]	[0]	402
180–190	152	11	[0]	[0]	163
195–200	345	34	[0]	[0]	379
260–275	63	93	[0]	[0]	156
282–290	185	179	[0]	[0]	364
320	23	50	[0]	[0]	73
360–380	215	233	[0]	[0]	448
420	41	34	[0]	[0]	75
457–470	201	184	[0]	[0]	385
505	29	35	[0]	[0]	64
560–585	[2]	40	47	[0]	89
595–612	[0]	70	152	[0]	222
700–740	[0]	7	219	[0]	226
750–755	[0]	2	62	[0]	64
840–890	[0]	[0]	227	38	265
897	[0]	[0]	37	[0]	37
980–1030	[0]	[0]	78	202	280
1040	[0]	[0]	28	13	41
1120–1182	[0]	[0]	[4]	310	314
1320–1377	[0]	[0]	[2]	267	269
1520–1610	[0]	[0]	[0]	216	216
1720–1770	[0]	[0]	[0]	123	123
1925–8120	[0]	[0]	11	82	93

DSs are E (egg and pharate first larval instar), L1 (egg hatching, first larval instar, and pharate second larval instar), L2 (first larval molt, second larval instar, and pharate third larval instar), and L3-f (second larval molt and feeding third larval instar).

Framed numerals (e.g., [0]) indicate that, for a mystery specimen in that DS, the p value is less than 0.05 for that ADH bracket.

2.5. Assessment of Coverage Rates

We assessed the statistical properties of the resulting confidence sets by cross-validation using each insect specimen as a singleton holdout sample. The DS of each of the insect specimens was compared to the ADH-wise frequency distributions of stages of the other (less the holdout) specimens, producing p values for each of the ADH ranges (24 as shown in Table 3). We tabulated the proportions of insect specimens from each (true) ADH bracket for which each (true or false) ADH bracket was rejected (the complement of the coverage rate), as shown in Fig. 1.

Computations were performed using SAS version 9.4 (SAS Institute Inc 2012).

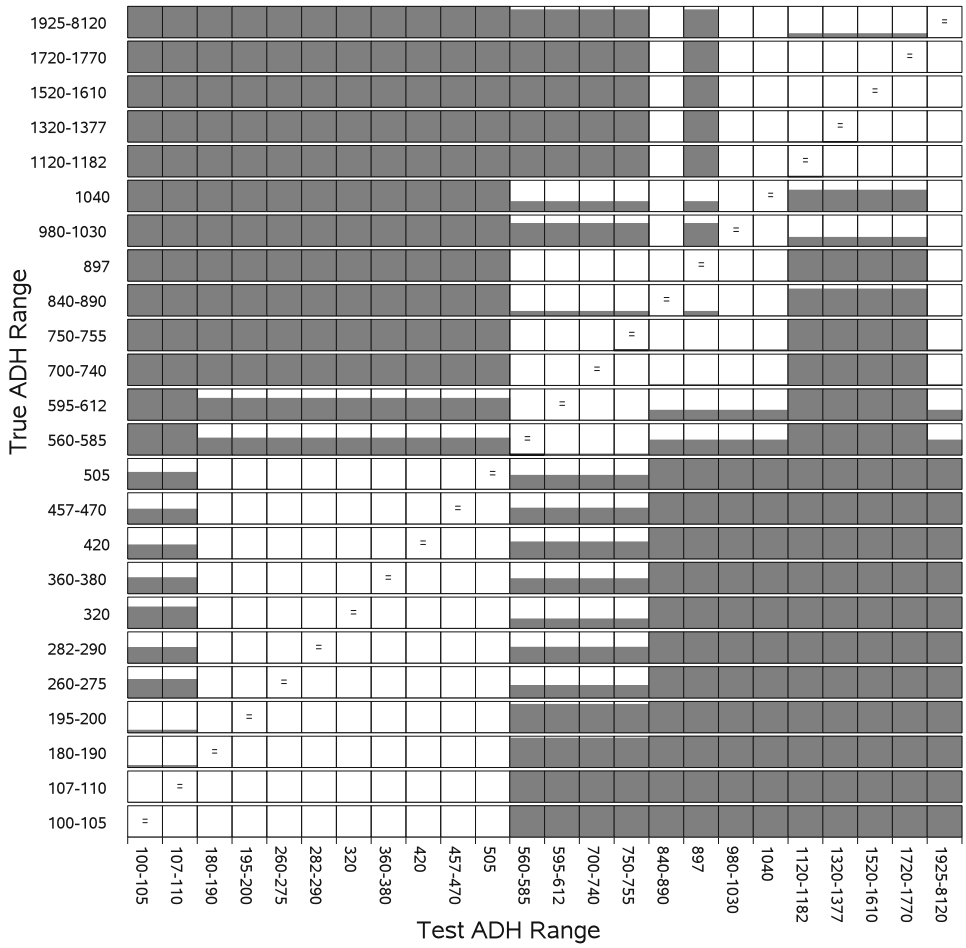


Figure 1. Results of cross-validation of rejection rates (1 minus coverage rates) of confidence sets on ADH from DS. Data are shown in Table 3. Each individual in the training data set was used as a holdout sample, and a confidence set on its ADH was constructed based on the remaining individuals. ADH bracket labels on both columns and rows are the same as in Table 3. Across each row, the *gray bar in each column* indicates the proportion of individuals in that ADH class for which the ADH class of that column was rejected. Diagonal cells, where True ADH = Test ADH, are indicated by "=" in the middle. A *black bar* within a diagonal cell (e.g., 1120–1182) indicates the rejection rate of the true ADH, which occurred for 10 of the 4902 *L. sericata* specimens. Zoom in to see detail.

3. RESULTS

3.1. Confidence Sets and Coverage Rates

Table 3 shows TD frequency distributions of the four DSs by ADH bracket. In each DS column, whether the p value comparing a MS of that DS to the frequency distribution of DSs within an ADH bracket is ≤ 0.05 is indicated by a frame around the frequency. Thus for an L1 MS, ADHs $\leq$110 degree hours or $\geq$840 degree hours are untenable. The 95% confidence set on ADH for such an insect specimen comprises the ADH brackets between 180–190 degree hours and 750–755 degree hours. In a similar fashion, the confidence set

for an egg (E) extends from the least bracket, 100–105 degree hours, through 505 degree hours (see Sect. 4).

Figure 1 portrays rejection rates within each true ADH bracket (horizontal row) for potential ADH brackets (columns), based on cross-validation in which each insect is a singleton holdout sample and p values compare its DS to distributions of DSs across all ADH brackets. Take row 505 ADH from Table 3 as an example: Of its 64 insects 29 were E and 35 were L1. The first ADH bracket, 100–105 degree hours, is not rejected for any of the 29 E insects, but it is for all of the 35 L1 insects. The rejection rate for the first ADH bracket among the 64 insects from 505 ADH is thus $35/64 = 55\%$ that is the height of the gray bar. As a second example, consider the 64 insects from ADH 750 to 755, with 2 in L1 and 62 in L2. When one of the two in L1 is held out, leaving 63 with one in L1, the p value for the holdout insect specimen (L1) for its true ADH bracket is less than 0.05 (see Table 4, where the n required for 5% significance when $x = 1$ is 51, which is less than the 63 left after taking out the one insect specimen), and so the 95% confidence set for this individual does not contain its true ADH. The same is true of the second L1 individual, and so the rejection rate for the true ADH bracket is $2/64 = 3\%$. This can be seen as the very short black bar in the 750–755 ADH diagonal cell.

Across a given row in Fig. 1, blank bars show that the column (vertical) ADH bracket was not rejected for any individual from the row (horizontal) ADH. All ADH brackets with blank bars, then, are included in the 95% confidence set for all individuals from the row ADH. For insect specimens from the row 260–275 ADH, for example, all confidence sets include ADH brackets 180–190 through 505, and $63/156 = 40\%$ include the first two brackets.

4. DISCUSSION

This illustration shows that the method of LaMotte and Wells (2000) can be used for statistical inference of ADH (and hence age and PMI_{min}) from DS. Application of this method for constructing confidence sets on ADH is made simple and accessible by using Table 3 in LaMotte and Wells (2000) or Table 4 in this paper. (Those tables are not identical. LaMotte and Wells (2000) included fewer categories (c) than shown in Table 4, but it gave minimum n for rejection at the 1, 5, and 10% levels of significance.) As assessed with cross-validation here, coverage rates (rejection rates) of true ADHs are greater (less) than the nominal 95% (5%), and they decrease (increase) for ADHs different from the true value.

Several disclaimers should be made concerning real applications. Although the probability statements underlying this method apply only to the age intervals in Table 3, we think conservative inferences can be made about most of the gaps between ADHs. Based on statistical inference, the only strong conclusion is to reject. Thus, the framed frequencies in Table 3 indicate that there is statistical evidence against the assertion that the MS came from that ADH bracket. Unframed frequencies, on the other hand, indicate lack of evidence to reject, so weak an indication as to be no conclusion at all. There is no statistical basis; then, to think that an L2 MS, for example, did not come from some ADH between 505 and 1120, exclusive of the endpoints. Indeed, without assuming a model to bridge the ADH

Table 4. Minimum training data sample size n for a significant difference (at the 5% level) between a single mystery specimen (MS) and training data (TD) comprising n individuals distributed over c categories.

	Minimum sample sizes for 5% significance										
	$x = 0$	1	2	3	4	5	6	7	8	9	10
$c = 2$	$n = 7$	17	28	39	52	64	77	91	104	118	132
3	15	34	55	78	102	128	153	180	207	235	263
4	22	51	83	117	153	191	230	269	310	351	393
5	29	67	110	156	204	254	306	359	413	468	524
6	37	84	137	195	255	318	382	448	516	584	654
7	44	101	165	234	306	381	458	538	619	701	785
8	52	118	192	272	357	444	535	627	722	818	915
9	59	135	220	311	408	508	611	717	825	934	1046
10	66	151	247	350	459	571	687	806	927	1051	1176
11	74	168	275	389	509	634	763	895	1030	1168	1307
12	81	185	302	428	560	698	840	985	1133	1284	1438
13	88	202	329	467	611	761	916	1074	1236	1401	1568
14	96	219	357	506	662	825	992	1164	1339	1518	1699
15	103	235	384	544	713	888	1068	1253	1442	1634	1829
16	110	252	412	583	764	951	1145	1343	1545	1751	1960

This is an extension of Table 3 in LaMotte and Wells (2000) for the 5% level of significance. The TD frequency in the MS category is x. For example, suppose $c = 4$ and the number of TD specimens sampled at a given, specific age is n. If the TD frequency in the MS's category is $x = 0$, then that age is rejected as being the age of the MS if $n \geq 22$; if $x = 1$, then that age is rejected if $n \geq 51$; and so on.

gaps, there is no statistical basis to make any inference about ADHs not represented in the training data.

There is a statistical basis to assert that an L2 MS did not come from an ADH between 457 and 470, for example. Further, although there is no direct statistical evidence to say that an L2 MS did not come from, say, ADH 480 (between the two brackets 457–470 and 505), if we surmise that the development process is reasonably smooth in that bracket, then the assertion that the MS came from ADH 480 would be rejected, too. Within reason and consideration, gaps between adjacent ADH brackets that were rejected could reasonably be rejected. Gaps between not-rejected ADH brackets must be regarded as not rejected, however, reflecting the lack of evidence for rejection.

The surfeit of interpretability in the gaps has implications for the design of the TD experiment. For example, the frequency of E individuals was 29 at 505 degree hours, dropping to 2 at the next observed ADH group, 560–585 degree hours. In hindsight, for our purposes brackets between about 195 and 470 degree hours were not necessary to sample, but more dense sampling around the age of egg eclosion would have helped to more precisely define the upper limits of the confidence set. Similarly, the upper boundary of stage L3-f is not distinguished well due to truncating ADH at 8120 degree hours.

Ideally, in each horizontal row of Fig. 1, we would like to see the gray bars squeeze in on the true ADH group, as shown by the squares along the diagonal, thus trimming down the possible values of ADH that cannot plausibly be eliminated. The power to reject values is limited by sample size, and it is the range of sample sizes (from 37 to 448) among the ADH brackets that accounts for much of the white space and the unevenness of the extent of the

gray bars. Overall, though, Fig. 1 shows that these interval estimates correspond well with true ADH. Most important is that they have a foundation in statistical methodology.

It is interesting to note that the rule "reject this ADH if the frequency among the TD of the MS stage is 0" gives results similar to those shown in Fig. 1. However, this rule would erroneously reject for any zeroes where the sample size was less than $n = 22$. It would fail to reject in those places in Table 3 where the framed frequencies are positive. Such examples would occur more frequently with greater sample sizes and more dense sampled ADHs. Finally, this ad hoc rule has no basis in probability, and no level of confidence could be defensibly attached to the results.

In several instances, there were significant differences (not shown here) among the frequency distributions from different age–temperature combinations at the same ADH. This indicates that there were effects on development rate not accounted for by ADH.

As noted already, multiple cohorts were observed at each age–temperature combination. For four cohorts of twenty specimens each, for example, there is the question, whether the underlying true distributions of stages are the same for all four. We addressed this question by comparing frequency distributions among cohorts within age–temperature combinations, and in several instances there were significant differences, but not accounting for multiple-testing inflation of the level of significance. That the cohorts are not homogeneous cannot be ruled out. The extent to which this impacts the p values used to construct the confidence sets on ADH is not known. We have begun to investigate this, and preliminary results indicate that the p values described in LaMotte and Wells (2000) are accurate under heterogeneity of this sort.

Anyone designing a carrion insect development experiment should consult the sample size thresholds in Table 3 in LaMotte and Wells (2000) or Table 4 in this paper. The greater the sample sizes and the finer the grid of ADHs, the narrower and more continuous the resulting confidence sets on age are likely to be. Sampling at different temperatures that produce the same ADHs would maximize sample size as well as provide data to address whether ADH adequately captures the joint effects of age and temperature. To the extent possible, age–temperature combinations where transitions between stages occur should be sampled.

ACKNOWLEDGEMENTS

This project was supported by Award Nos. 2013-DN-BX-K042 (LRL and JDW) and 2010-DN-BX-K231 (LGH), awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

[Received November 2016. Accepted April 2016. Published Online April 2017.]

REFERENCES

Adams, Z.J., Hall, M.J.R. (2003). Methods used for the killing and preservation of blowfly larvae, and their effect on post-mortem larval length. Forensic Science International 138:50-61.

Amendt, J., Campobasso, C.P., Gaudry, E., Reiter, C., LeBlanc, H.N., Hall, M.R.J. (2007). Best practice in forensic entomology - standards and guidelines. International Journal of Legal Medicine 121:90-104.

Baque, M., Filmann, N., Verhoff, M.A., Amendt, J. (2015). Establishment of developmental charts for the larvae of the blow fly *Calliphora vicina* using quantile regression. Forensic Science International 248:1-9.

Catts, E.P. (1990). Analyzing entomological data. In: Catts, E.P., Haskell, N.H. (Eds.), Entomology & Death: A Procedural Guide, pp. 24-35. Joyce's Print Shop, Clemson, South Carolina.

Dadour, I.R., Cook, D.F., Wirth, N. (2001). Rate of development of *Hydrotaea rostrata* under summer and winter (cyclic and constant) temperature regimes. Medical and Veterinary Entomology 15:177-182.

Gaudry, E., Myskowiak, J-B., Chauvet, B., Pasquerault, T., Lefebvre, F., Malgorn, Y. (2001). Activity of the forensic entomology department of the French Gendarmerie. Forensic Science International 120:68-71.

Goff, M.L. (1993). Estimation of postmortem interval using arthropod development and succession patterns. Forensic Science Review 5:81-94.

Huntington, T., Higley, L. (2008). Collection and analysis of climatological data. In: Haskell, N.H., Williams, R.E. (Eds.), Entomology & Death: A Procedural Guide. 2nd edition, pp. 144-159. East Park Printing, Clemson, South Carolina.

Ieno, E.N., Amendt, J., Fremdt, H., Saveliev, A.A., Zuur, A.F. (2010). Analyzing forensic entomology data using additive mixed effects modeling. In: Amendt, J., Goff, M.L., Campobasso, C.P., Grassberger, M. (Eds.), Current Concepts in Forensic Entomology, pp.139-163. Springer, Heidelberg.

LaMotte, L.R., Wells, J.D. (2000). P-values for postmortem intervals from arthropod succession data. Journal of Agricultural, Biological and Environmental Statistics 5:37-47.

—— (2015). Inverse prediction for heteroscedastic response using mixed models software. Communications in Statistics – Simulation and Computation, accepted author version posted online 21 Nov. 2015. DOI:10.1080/03610918.2015.1118508.

—— (2016). Inverse prediction for multivariate mixed models with standard software. Statistical Papers 57(4):929-938. DOI:10.1007/s00362-016-0815-2.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses, Second Edition*. John Wiley & Sons, Inc.

Midgley, J.M., Villet, M.H. (2009). Effect of the killing method on post-mortem change in length of larvae of *Thanatophilus micans* (Fabricius 1794) (Coleoptera: Silphidae) stored in 70% ethanol. International Journal of Legal Medicine 123:103-108.

National Research Council (2009). Strengthening Forensic Science in the United States, A Path Forward. National Academies Press, Washington.

Roe, A., Higley, L.G. (2015). Development modeling of Luciliasericata (Diptera: Calliphoridae). PeerJ 3:e803. DOI:10.7717/peerj.803.

SAS Institute Inc. (2012). SAS 9.4. Cary, N. C.

Tantawi, T.I., Greenberg, B. (1993). The effect of killing and preservative solutions on estimates of maggot age in forensic cases. Journal of Forensic Sciences 38:702-707.

Villet, M.H., Richards, C.S., Midgley, J.M. (2010). Contemporary precision, bias and accuracy of minimum post-mortem intervals estimated using development of carrion-feeding insects. In: Amendt, J., Goff, M.L., Campobasso, C.P., Grassberger, M. (Eds.), Current Concepts in Forensic Entomology, pp. 109-137. Springer, Heidelberg.

Wall, R., French, N., Morgan, K.L. (1992). Effects of temperature on the development and abundance of the sheep blowfly *Lucilia sericata* (Diptera: Calliphoridae). Bulletin of Entomological Research 82:125-131.

Wells, J.D. , LaMotte, L.R. (1995). Estimating maggot age from weight using inverse prediction. Journal of Forensic Sciences 40:585-590.

—— (2001). Estimating the postmortem interval. In: Byrd, J.H., Castner, J.L. (Eds.), Forensic Entomology, The Utility of Arthropods in Legal Investigations, 1st Edition, pp. 367-388. CRC Press, Boca Raton, Florida.

Wells, J.D., Lecheta, M.C., Moura, M.O., LaMotte, L.R. (2015). An evaluation of sampling methods used to produce insect growth models for postmortem interval estimation. International Journal of Legal Medicine 129:405-410.