Chapter II

Statistical Design and Analysis of Date Palm Insect Pest Management Experiments

Murari Singh[1], Salim Al-Khatri[2], Khaled El-Shamaa[1], and Abdoul Aziz Niane[3]

[1]ICARDA, Amman, Jordan

[2]Plant Protection Research Centre, Ministry of Agriculture and Fisheries, Muscat, Sultanate of Oman.

[3]ICARDA, Beirut, Lebanon

Email: M.Singh@cgiar.org

1. Introduction

This chapter discusses selected experimental design and data analyses in the context of date palm insect experiments. This chapter illustrates the number of infested fruit, an analysis of repeated measurements, and estimates the number of juvenile nematodes of two species and three sizes found in a date palm species with real data. It presents an analysis of dose-binary response data with an aim to estimate the lethal dose and it provides a World Wide Web link for computation[1]. In a date palm experiment with insect pests, one may be interested in controlling the insect population or the effects on fruit damage by applying a number of newly developed chemical or bio-control insecticides and organic preparations. A detailed and systematic description of establishing date palm in a suitable environment/land is presented by Zaid and Botes (2002) and Zaid et al. (2002). Multiple date palm trees of various varieties with similar planting date are grown such that trees of the same age are available as effective controls of insect pests, including the application of insecticides. One or many insecticides may be applied on infested date palms. Treated palms are observed by recording insect counts or yields over several days within a meaningful period of time. The general objectives in these situations are to estimate and to compare the effects of the insecticides or control measures. Integrated Pest Management (IPM) experiments on date palm may involve a wide range of objectives. Some examples include the study of the following factors: effect of pesticides on insect mortality rates and yield on a date palm variety, surveys to identify the locations with high prevalence of various date palm insect pests, associations between the pest infestation and clustering of locations for similar pest incidences, estimation of the peak period for infestation of date palm pests, modeling infested plants in order to study the spatial and temporal distribution of infestation rates.

We discuss the data analysis of the following three experiments.

Study 1: Consider a date palm experiment with a view to control the effect of *Batrachedra amydraula* Meyrick on fruit infestation using 5 insecticides on the branches. The

[1] http://geoagro.icarda.org/bss/shinyapps/ld50

experimental design was a completely randomized design with 6 treatments (including water as a control) each with six trees as replications. Thus, the insecticides were randomly applied to the trees. Fruits on three branches on each tree were examined for infested and healthy (un-infested) fruits. The observations were taken on a weekly basis. The objective was to examine and to compare the effectiveness of the insecticides in controlling the fruit infestation.

Study 2: In another study on entomopathogenic nematodes, *Steinernema feltiae* and *Heterorhabditis bacteriophora* were counted on date palms over a period of time. The nematodes of each species varied in weight (or size) and were grouped as small, medium, or large. Each species and each size group of nematodes were counted, for juveniles, on each of the five randomly chosen trees for 4 – 37 days with an interval of 3-4 days. The objective was to examine any association (interaction) between species and size of the nematodes for the infective juvenile numbers as well as their dynamics over time.

Study 3: Dose – response relationship to control small grain storage insects. Fifteen samples of seed and grains from wheat and barley infested with *Rhizopertha dominica* (Fabricius) were collected from storage facilities in the North of Syria. Three-week-old populations of *R. dominica* were reared from the samples collected and exposed to variable doses of Phosphine (PH3), including a discriminating dose for this insect species, which is 0.03 mg/l PH3 for 20 hrs. At the end of this fixed exposure time (20 hrs), the insects were incubated under optimal environmental growing and reproduction conditions for *R. dominica* at 70% RH and 25 ˚C for 14 days. The insect populations were then sorted into two categories: responded (killed) and non-responded (survived).

2. Experimental designs

Some basic concepts and commonly used experimental designs in date palm pest experiments are described below.

2.1. Elements of experimental designs

Treatments refer to the different factors or procedures intended to create variation in a response (responses) in an experiment, e.g., insecticides.

An experimental unit is the smallest size of the experimental material to which the treatment is applied, such that any two units may receive different treatments. For example, a palm tree is an experimental unit to which an insecticide is applied while a neighboring palm tree may be applied a different insecticide. If instead of one palm tree, one has sets of 5 trees grown together and the same treatment is applied to the set of 5 trees, then the set of 5 trees is an experimental unit, provided any such sets may receive different treatments.

Experimental Material is the collection of all experimental units for the chosen experiment. For example, all the palm trees used for the experiment.

An experimental design is used to estimate and to compare treatment effects on a response variable (e.g., fruit yield, number of infested fruits) with a high degree of precision. Even if the same treatment has been applied on a number of homogeneous experimental units, a variation in a response is observed and may have arisen due to uncontrolled causes. This is called experimental error variation and is essential to obtain the precision of an estimate of the effect or difference of means. It is desirable to have a good experimental design which estimates treatment effects/comparisons from any systematic variation in the experimental material, high precision, valid comparisons with measurable uncertainty and generalizable over a wide range of conditions or environments.

2.2. Fisher's principles of experimentation

$$r = \frac{\theta^2 t^2}{\varepsilon^2}$$

Where

θ = coefficient of variation ($\frac{\sigma}{\mu}$),

t= critical value of t- distribution (r-1 df) and approximated at 2 for 5% level of significance,

ε= maximum error set, $\left|\frac{\bar{x}-\mu}{\mu}\right|$, where $\bar{x}$ is sample mean expected from r replications, and μ is the population mean (unknown).

Some standard texts on basics of experimental designs and analysis include Cochran and Cox (1957), Gomez and Gomez (1984), Hinkelmann and Kempthorne (2005), and a review by Singh and El-Shamaa (2015).

When designing an experiment for IPM on date palm, the following situations may arise:

Situation 1: the experimental material is fully homogeneous.

If the experimental material is homogeneous, e.g., all palm trees are of same genotype, same age, and grown and cared in the same environment, one may randomly apply the experimental treatments with the same or a variable number of replications. Such a design is called Completely Randomized Design (CRD). In this situation, the total variability is partitioned through a mechanism called analysis of variance (ANOVA) into the sources of variation due to treatment and experimental error.

Situation 2: The experimental material is partly homogeneous.

If the experimental material is partly homogeneous, Local Control or Reduction of Error is done by accounting for any systematic variation in the experimental material at either the design stage or at the analysis stage or both. One example of control is practiced by forming homogeneous blocks or groups of experimental units. Such an experimental design is called a randomized complete block design (RCBD). The treatments are randomly allotted to the units within each block. The sources of variation to account for the total variation are

blocks, treatments, and experimental error. Examples of blocking may be age of the trees, location of the trees, etc. After block variation has been accounted for, RCBD reduces the experimental error relative to CRD.

3. Analysis of data from designed experiments

The standard analysis of data from a design is based on expressing the response as a linear model in terms of effects of various factors, such as blocks and treatment, and an uncontrolled (experimental) error. The analysis of variance (ANOVA) is a method which partitions the total variation in the response into the components (sources of variation) in the above model. The following assumptions are validated before drawing inferences on the treatments: additivity of factors effects, constancy of error variance, normality of experimental errors, and independence of experimental errors A statistical software is used to carry out the computations. We here consider two specific cases of data analysis.

3.1. Analysis of data with repeated measures

In the context of Study 1 and in order to evaluate the effects of the five insecticides and a control on fruit damage, a completely randomized design with 6 treatments (including water as a control), each with six trees as replications, is implemented. Over 5 weeks, the numbers of infested fruits were observed for three individual branches in each tree:

Treatment	Tree	Branch	InfFruits0	InfFruits1	InfFruits2	InfFruits3	InfFruits4
Control	1	1	7	2	18	36	1
Control	1	2	4	4	5	18	5
Control	1	3	3	13	11	7	3
Control	2	1	6	6	7	7	1
Control	2	2	6	5	6	10	2
.							
.							
.							
Insecticide A	2	1	8	2	2	6	8
Insecticide A	2	2	8	7	13	16	7
Insecticide A	2	3	8	3	8	4	1
Insecticide A	3	1	12	0	1	5	7
Insecticide A	3	2	1	2	1	7	6
Insecticide A	3	3	2	2	2	9	6
Insecticide A	4	1	13	0	0	8	3
Insecticide A	4	2	17	0	0	4	5
Insecticide A	4	3	13	0	4	1	10
Insecticide A	5	1	0	1	2	0	0
Insecticide A	5	2	3	1	4	1	3

The observations on the same branch over the weeks are correlated. Furthermore, the number of fruits in the observed range may require square-root transformation before analysis using repeated measures method to test significance of insecticide and week interaction and estimate their effects. The following Genstat directives were used in the analysis:

```
AREPMEASURES [PRINT=epsilon, test; APRINT=aovtable,
information,mean,%cv; TREATMENT=Treatment;\
BLOCK=Tree.Treatment/Branch; FPROB=yes; PSE=diff, lsd,
means; LSDLEVEL=5;\
TIMEPOINTS=!(0,1,2,3,4); FACT=9]SqrtInfFruits0, \
SqrtInfFruits1,SqrtInfFruits2,SqrtInfFruits3,SqrtInfFru
its4
```

where Tree, Branch, Treatment and Week are factors standing for the date palm tree (1-6), branch (1-3), insecticides (A-D, Control) and weeks (0-4). The square-root transformed values of the number of infected fruits during weeks 0 to 4 are SqrtInfFruits0, SqrtInfFruits1, SqrtInfFruits2, SqrtInfFruits3, SqrtInfFruits4, respectively.

Partial output:

Box's tests for symmetry of the covariance matrix
Chi-square 24.06 on 13 degrees of freedom: probability 0.031
F-test 1.85 on 13 and 59480 degrees of freedom: probability 0.031
Greenhouse-Geisser epsilon
Epsilon: 0.9346
Analysis of variance
Variate: SqrtInfFruits0,SqrtInfFruits1,SqrtInfFruits2,SqrtInfFruits3,SqrtInfFruits4

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Tree.Treatment stratum					
Treatment	5	251.6054	50.3211	12.59	<.001
Residual	30	119.9434	3.9981	4.26	
Tree.Treatment.Branch stratum					
	72	67.6233	0.9392	1.17	
Tree.Treatment.Branch.Time stratum					
d.f. correction factor 0.9346					
Time	4	102.7442	25.6861	31.95	<.001
Treatment.Time	20	118.0055	5.9003	7.34	<.001
Residual	408	327.9775	0.8039		
Total	539	987.8993			

(d.f. are multiplied by the correction factors before calculating F probabilities)
Tables of means
Variate: SqrtInfFruits0,SqrtInfFruits1,SqrtInfFruits2,SqrtInfFruits3,SqrtInfFruits4
Grand mean 1.781

Treatment

Control	Inscticide_A	Insecticide_B	Insecticide_C	Insecticide_D	Insecticide_E
2.299	2.058	0.762	2.717	1.800	1.049

Time	0	1	2	3	4
	2.450	1.281	1.632	2.111	1.430

Time	0	1	2	3	4
Treatment					
Control		2.234	1.721	2.464	3.352	1.723
Inscticide_A		2.902	1.314	1.722	2.303	2.046
Insecticide_B		1.920	0.780	0.596	0.458	0.056
Insecticide_C		2.328	1.960	3.350	3.430	2.516
Insecticide_D		2.752	0.994	1.105	2.405	1.746
Insecticide_E		2.563	0.918	0.556	0.715	0.490

Standard errors of means

Table	Treatment	Time	Treatment Time
rep.	90	108	18
e.s.e.	0.2108	0.0863	0.2831
d.f.	30	381.32	92.93

Except when comparing means with the same level(s) of

Treatment	0.2113
d.f.	381.32

Correction factors have been applied to residual d.f.(see analysis-of-variance table for details)

Least significant differences of means (5% level)

Table	Treatment	Time	Treatment Time
rep.	90	108	18
l.s.d.	0.6087	0.2428	0.8045
d.f.	30	381.32	92.93

Except when comparing means with the same level(s) of

Treatment	0.5947
d.f.	381.32

Correction factors have been applied to residual d.f.(see analysis-of-variance table for details)

In the above ANOVA table,

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Treatment	5	251.6054	50.3211	12.59	<.001
and Treatment.Time	20	118.0055	5.9003	7.34	<.001

Interactions between the treatments (insecticides and control) and week are statistically significant. Further, significant differences in overall means (main effects) of the treatments are observed as p-values (F-probability) and are very low, $P < 0.001$. Using back-transformation to the original scale of number of fruits, we need to square the means based on square-roots. However, the standard errors associated with these squared values will be different for different means. In order to compare treatments for the means, we keep the SqrtInfFruits0-4 variable means with their common standard errors.

The significant interaction indicates the choice for selecting the insecticide, which would be more effective in a desired week. Furthermore, we notice considerable variability in the means under week 0, particularly for insecticides A and B. Insecticide B controls the fruit damage significantly, ($P < 0.05$) most effectively (i.e., in relation to the control) during weeks 1, 3 and 4.

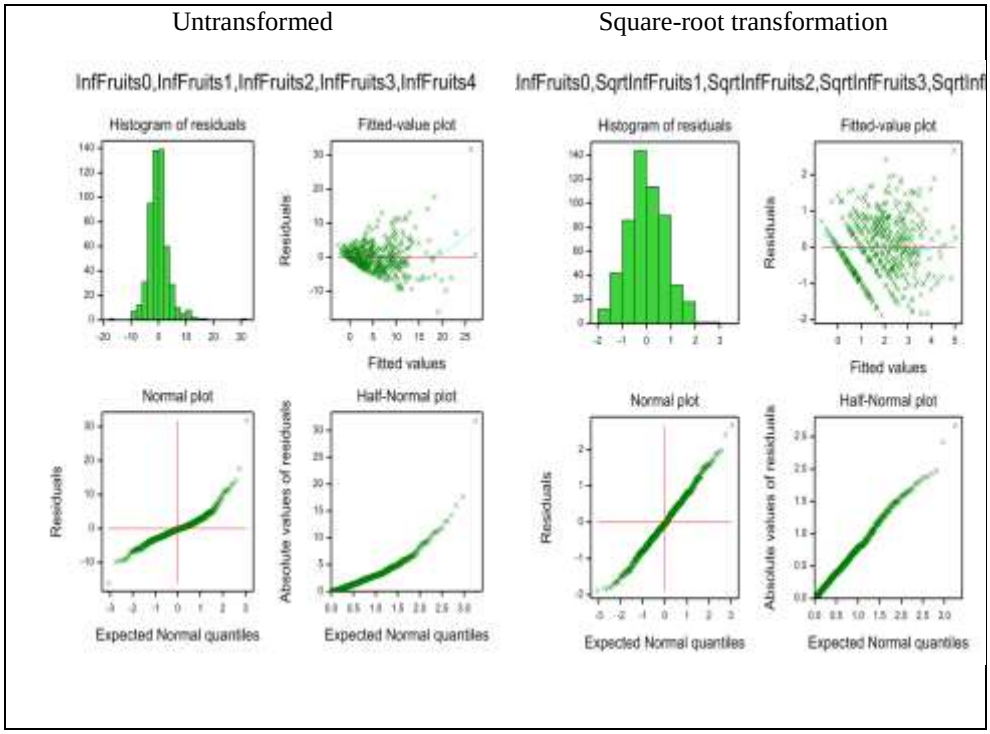

Fig.1. Residual plots for infested fruit number (left panel) and its square-root (right panel): Histogram of residuals (left upper row), residual versus fitted values (right upper row), quantile plots of residuals (left lower row) and quantile plot of absolute values of residuals (right lower row).

Study 2: In the study on nematodes of two species and different sizes, the number of infective juveniles found on a species of date palm was observed. A subset of data on the number of counts over a period of 37 days is shown below:

Nematode	Galleria Weight	Replicate	Infective Juveniles
S.feltiae	Small	1	75032
S.feltiae	Small	2	180615
S.feltiae	Small	3	51900
S.feltiae	Small	4	160500
S.feltiae	Small	5	137100
S.feltiae	Small	6	7392
S.feltiae	Small	7	94206
S.feltiae	Small	8	15428
S.feltiae	Small	9	90400
S.feltiae	Small	10	75851
S.feltiae	Medium	1	140200
S.feltiae	Medium	2	64384
S.feltiae	Medium	3	262100
S.feltiae	Medium	4	129500
S.feltiae	Medium	5	127600
S.feltiae	Medium	6	69858
S.feltiae	Medium	7	57165

The association of the average number of infective juveniles and the species and sizes can be examined using an analysis of variance. However, we also need to adjust for the heterogeneity of variances that may arise with species and size. These numbers were log-transformed and their means and variances are as follows:

Nematode	Weight	N	Mean(IJs)	Var(IJs)	Mean(LogIJs)	Var(LogIJs)
H.bacterip	Large	10	436202	2.43E+10	12.86	0.4293
H.bacterip	Medium	10	350104	5.34E+09	12.74	0.0499
H.bacterip	Small	10	244756	4.72E+09	12.37	0.0864
S.feltiae	Large	10	147778	1.02E+09	11.88	0.0427
S.feltiae	Medium	10	115029	3.53E+09	11.55	0.205
S.feltiae	Small	10	88842	3.3E+09	11.07	1.0613
p-value for homogeneity of variances				P < 0.001		P < 0.001

The resulting boxplots were as follows:

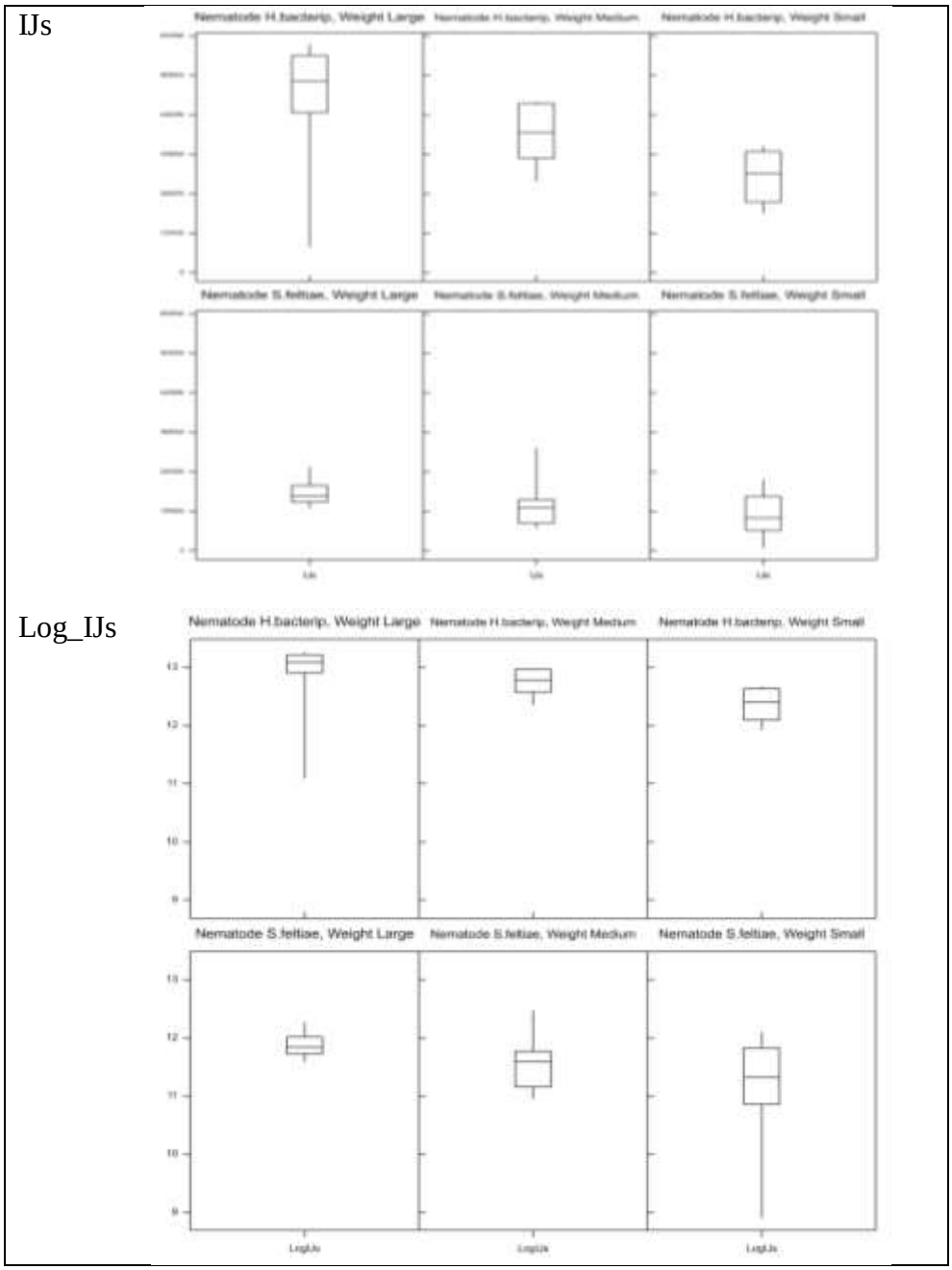

Fig.2. Boxplots of the number of infective juveniles (left panel) and its logarithmic transformation (right panel) for the two species (rows) and three sizes (columns) of the nematodes.

Ignoring the variance heterogeneity, the ANOVA gives an assessment of the association of the mean number of infective juveniles with species and size, but no interaction, at the log-transformed values. We also need to note that the means here are per replicate basis.

```
CALCULATE LogIJs=log(IJs)
BLOCK "No Blocking"
TREATMENTS Nematode*Weight
COVARIATE "No Covariate"
ANOVA [PRINT=aovtable,information,means; FACT=32;
CONTRASTS=7; PCONTRASTS=7; FPROB=yes;   PSE=diff,lsd,means;
LSDLEVEL=5] LogIJs
Analysis of variance
```

Variate: LogIJs

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Nematode	1	20.0852	20.0852	64.28	<.001
Weight	2	4.4231	2.2115	7.08	0.002
Nematode.Weight	2	0.2779	0.1389	0.44	0.643
Residual	54	16.8726	0.3125		
Total	59	41.6587			

Tables of means
Variate: LogIJs
Grand mean 12.080

Nematode	H.bacterip	S.feltiae
	12.658	11.501

Weight	Large	Medium	Small
	12.372	12.149	11.718

Nematode Weight	Large	Medium	Small
H.bacterip	12.860	12.745	12.370
S.feltiae	11.884	11.554	11.066

Standard errors of means

Table	Nematode	Weight	Nematode Weight
rep.	30	20	10
d.f.	54	54	54
e.s.e.	0.1021	0.1250	0.1768

Least significant differences of means (5% level)

Table	Nematode	Weight	Nematode Weight
rep.	30	20	10
d.f.	54	54	54
l.s.d.	0.2894	0.3544	0.5012

```
 DELETE [REDEFINE=yes] _mean, _rep, _var, _resid, _rdf,
_scode
AKEEP [FACTORIAL=9] Nematode.Weight; MEAN=_mean; REP=_rep;
VARIANCE=_var; RTERM=_resid;   STATUS=_scode
 AKEEP [FACTORIAL=9] #_resid; DF=_rdf
 AMCOMPARISON [PRINT=letter; METHOD=bonferroni;
DIRECTION=descending; PROB=0.05; FACTORIAL=9]
Nematode.Weight
```

Bonferroni test
Nematode.Weight
Comparison-wise error rate = 0.0033

		Mean
H.bacterip	Large	12.86 a
H.bacterip	Medium	12.74 a
H.bacterip	Small	12.37 ab
S.feltiae	Large	11.88 bc
S.feltiae	Medium	11.55 cd
S.feltiae	Small	11.07 d

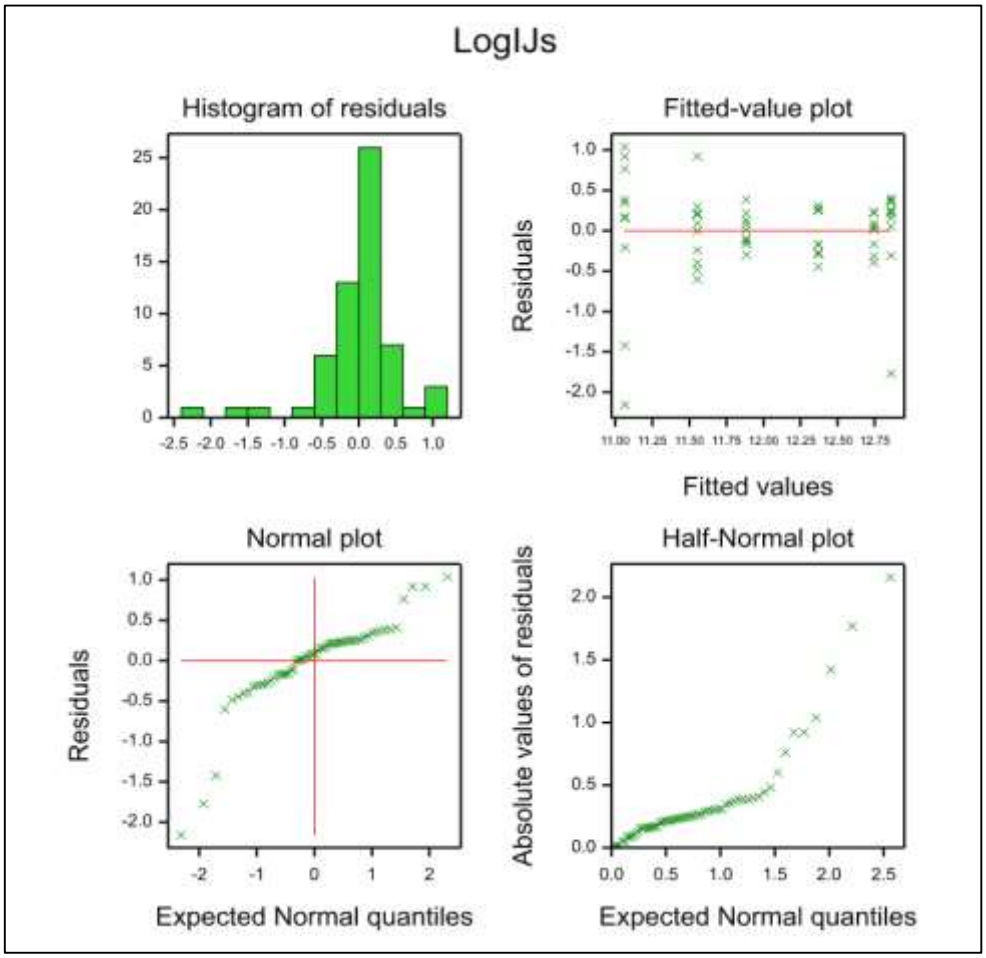

Fig.3. Residual plots of the logarithm of the number of infective juveniles: Histogram of residuals (left upper row), residual versus fitted values (right upper row), quantile plots of residuals (left lower row) and quantile plot of absolute values of residuals (right lower row)

Heterogeneity in the variances can be accounted using a weight vector Wet= (number of observation/variance) for the mean for combinations of species and size of the nematodes. In this case, we can use the residual maximum likelihood (REML). In the following output, the symbols Nematode0, Weight0, and yMn are factors for species, size, and variate for mean of logarithms of number of IJs, respectively. The vector Wet is the inverse of the variances of the means. These all are of length 6.

```
104              VComp[Fixed=Nematode0       +       Weight0]
105  Reml[Weight=Wet] yMn
REML variance components analysis
Response variate: yMn
Fixed model:      Constant + Nematode0 + Weight0
Number of units: 6
Weights variate:  Wet

Residual term has been added to model
Sparse algorithm with AI optimisation
 Residual variance model
 Term      Model (order)  Parameter          Estimate s.e.
Residual Identity  Sigma2      0.452 0.4517

Tests for fixed effects
Sequentially adding terms to fixed model
Fixed term        Wald statistic    n.d.f.    F statistic       d.d.f.    F pr
Nematode0         235.26              1         235.26           2.0      0.004
Weight0           45.23               2          22.61           2.0      0.042

Dropping individual terms from full fixed model
Fixed term        Wald statistic    n.d.f.    F statistic       d.d.f.    F pr
Nematode0         198.60              1         198.60           2.0      0.005
Weight0           45.23               2          22.61           2.0      0.042

Table of predicted means for Nematode0
          Nematode0        H.bacterip        S.feltiae
                             12.70 11.56
Standard errors
Average:          0.04956
Maximum:          0.05658
Minimum:          0.04254

Table of predicted means for Weight0
          Weight0          Large        Medium       Small
                           12.44         12.16        11.79
Standard errors
 Average:          0.05725
Maximum:          0.06920
Minimum:          0.04917
```

Note that the p-values accounting for the heterogeneity of variances have changed from the respective values in the ANOVA table based on homogeneous variances. However,

under both analysis scenarios, significant differences were observed between the species and the weight sizes. The number of IJ can be obtained using back-transformation. Using the REML analysis, the IJ for H.bacterip is 327748 [=exp(12.7)] and the IJ for S.feltiae is 104820. The IJs for the three body sizes, Large, Medium and Small, are 252711,190995, and 131927, respectively.

3.2. Estimation of doses for binary responses

Here, we present a more specific IPM situation, controlling the insect pests which cause damage to the fruits. Studying the relationship between dose and dichotomous (dead/alive, germinated/dormant, diseased/healthy) response is helpful in estimating the minimum lethal doses that will cause a desired response, for example, the dose which kills 50% of the insects.

To further our understanding by modeling the dose-response relationship, consider that a number of units (n) are exposed to a given dose (x) and suppose that m units responded. The response rate $p = m/n$ is expressed in terms of dose x. Several models could be used for the underlying mechanism of the response. One such model is cumulative and a sigmoid curve could be used to model a cumulative response rate.

Cumulative probability of response:

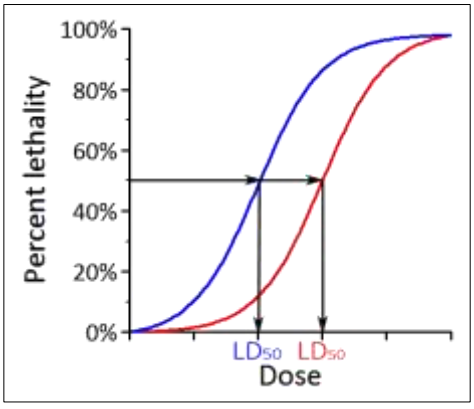

Fig. 4. A hypothetical model for mortality rate and dose relationship

Thus, a model needs to be found to satisfactorily describe the dose-response relationship. We discuss two such models.

Probit analysis:

The binary response is not a normal random variable. A normal random variable often forms the basis of many statistical analyses. However, one way is to contemplate the existence of a tolerance variable, say with value denoted by ζ at dose x. Such a variable is latent and

might be (or is assumed to be) normally distributed with unknown mean μ and variance σ^2. In this case, one can model the m responses out of n at dose x as

Probability (response at dose x) = $m/n = p$ = Cumulative normal(ζ ; μ , σ^2)= $\Pr(Tolerance \le \zeta) = \Phi((\zeta - \mu)/\sigma)$.

Inverting the above equation to write in terms of normal deviate:

$$\Phi^{-1}(p) = (\zeta - \mu)/\sigma$$

Or, the tolerance $\zeta = \mu + \sigma\Phi^{-1}(p)$ is assumed linear in dose x. Or, $\mu + \sigma\Phi^{-1}(p) = \alpha + \beta x$

Thus, we have the model in observed cumulative probability

$$\Phi^{-1}(p) = [(\alpha - \mu)/\sigma] + [\beta/\sigma]x \quad \text{Or,} \quad \Phi^{-1}(p) = \alpha^* + \beta^* x$$

which is a linear function of dose x. In probit analysis, we fit this function. Once $\alpha^* and \beta^*$ are estimated, we can estimate x for a given response rate, say 50%. Such a dose is called the lethal dose LD50%, where if response stands for dead (out of total alive).

$$\Phi(\alpha^* + \beta^* x) = p = 0.5 \text{. Or, } \alpha^* + \beta^* x = \Phi^{-1}(p) = \Phi^{-1}(0.5) = 0$$

Thus, $LD(50\%) = x = -\alpha^*/\beta^*$

A detailed discussion of probit analysis is available in Finney (1952).

Logit model:

Cumulative probability can also be modeled as a logit function:

A logit function of p (0< p <1), described as “log of odds ratio”, is defined as

Logit(p)=log($p/(1-p)$. It arises from a logistic function in x given as follows:

$p = 1/[1 + e^{-(\alpha+\beta x)}]$, or, $p/(1-p) = 1/e^{-(\alpha+\beta x)}$, or, $\ln(p/(1-p)) = \alpha + \beta x$, a linear function.

Using Generalized Linear model fitting programs (VSN International 2015), we can estimate α and β as well as the dose x at a given response, e.g.. 50%.

Study 3: Laboratory fumigation tests were conducted to determine the toxicity of different Phosphine dosages on storage pests (*Rhizopertha dominica*) using Probit analysis. The experimental design was a completely randomized design with three replicates of 50

insects, each exposed to seven dosages, 0.0, 2.4, 4.0, 8.0, 17.0, 21.0 and 30.0 mcg/l PH3, covering the anticipated full range of mortality including a control.

Dose	Replicate	Insects tested	Insects killed
0.0	1	51	0
0.0	2	49	2
0.0	3	49	2
2.4	1	52	2
2.4	2	50	1
2.4	3	50	2
4.0	1	51	8
4.0	2	51	7
4.0	3	52	7
8.0	1	50	14
8.0	2	50	45

Dose	Replicate	Insects tested	Insects killed
8.0	3	49	40
17.0	1	50	48
17.0	2	52	48
17.0	3	53	48
21.0	1	50	48
21.0	2	49	42
21.0	3	50	46
30.0	1	51	50
30.0	2	50	48
30.0	3	52	50

An Analysis Tool: In order to estimate the lethal dose for a specified mortality rate (e.g., 50%), one may use the following online tool: http://geoagro.icarda.org/bss/shinyapps/ld50 (El-Shamaa, 2017). This online tool was built using the R language and the Shiny framework for web applications; it fits a Generalized Linear Model (GLM) assuming that the error distribution is binomial; and it can apply three different link functions for probability transformations (Probit, Logit, and Complementary log-log). The user can define the effective (or lethal) dose/concentration and the level of confidence interval using interactive sliders in the left bar. The above web application URL can be accessed using your favorite browser. No statistical software is needed on your computer, simply upload your data and start work on your analysis online.

For this web application, data should be in the Excel file format, listed in the first sheet using column wise style starting from the A1 cell, where the first row contains the column labels. Required inputs include three columns:

- Number of subjects (e.g. total number of insects),
- Responded number (e.g. insects got killed), and
- Explanatory variate (e.g. dose).

An extra optional column refers to a grouping factor whose levels/labels may denote, for example, different pesticides.

Clicking the “Browse” button opens a normal file pop-up window. Select your Excel data file and then click “Open”. It will take a few moments (depending on your file size and your Internet speed) until the “Upload complete” message appears at the progress bar just below the “Browse” button. The data will now be listed in the main body of the “Input Data” tab (see Fig LD50Shiny-1.png).

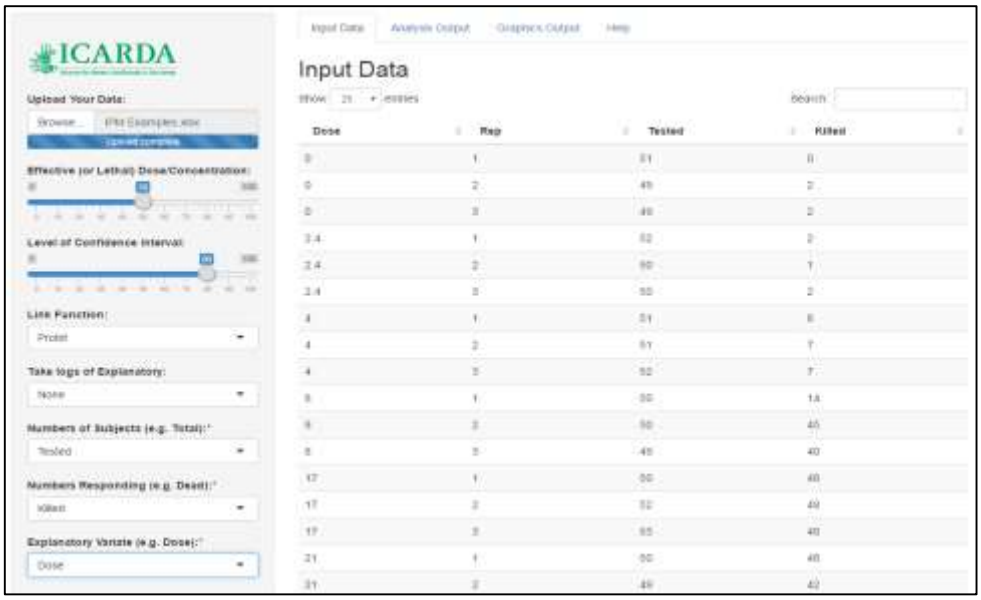

Fig. 5. A screen shot from an online tool at ICARDA: LD50Shiny-1

Analysis parameters include:

Effective (or lethal) dose/concentration (default is 50% and accepted value is in the range [0 – 100]).

Level of confidence interval (default is 80% and accepted value is in the range [0 – 100]).

Link function (default is “Probit” and available options includes also “Logit” and “Complementary log-log”).

Take logs of explanatory (default is “None” and available transformations includes log base 10 and log base e).

Once the data is uploaded, assign the required input columns to the subjects, responded, and explanatory. The web application provides a combo box for each input listing all the columns in your Excel data file in order to select the one associated with each input.

Switch to the “Analysis Output” tab to find out estimates of LDs. If there is a grouping factor, e.g., different types of pesticides, the analysis will be performed for each grouping level separately and provide standard errors, fitted model parameters, and summary statistics (see Fig LD50Shiny-2.png).

Further output can be obtained from the “Graphics Output” tab, which draws a graph of the fitted model showing the relationship of the response with the explanatory variable, including the confidence interval and LD cutting line for selected effective dose/concentration (see Fig. LD50Shiny-3). The user can download the resulting graph in high resolution (e.g., for publications) by clicking on the related button at the top of the graph. The contents of both “Analysis Output” and “Graphics Output” tabs are dynamic; in other words, if you change any of the analysis parameters in the left bar the results will affect and update instantly.

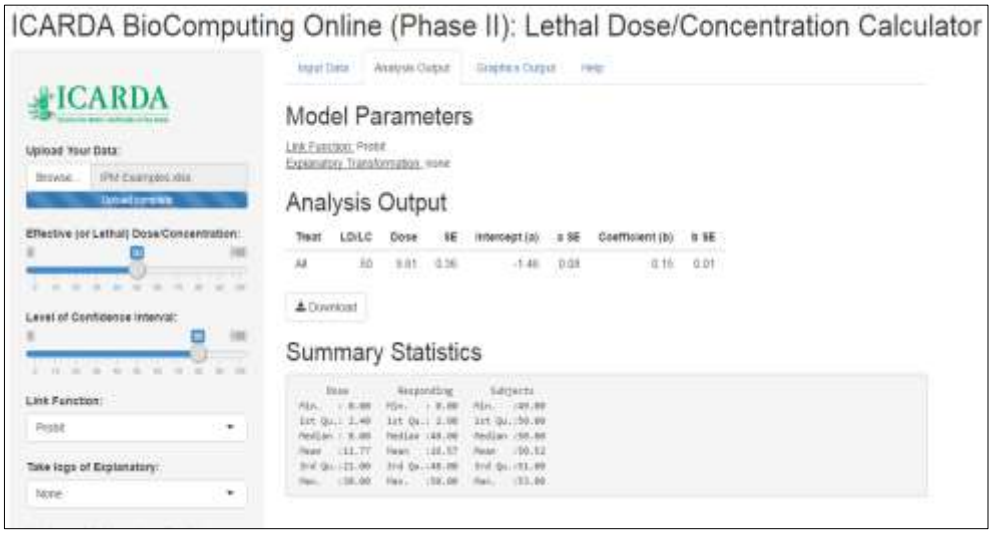

Fig.6. A screen shot from an online tool at ICARDA: LD50Shiny-2

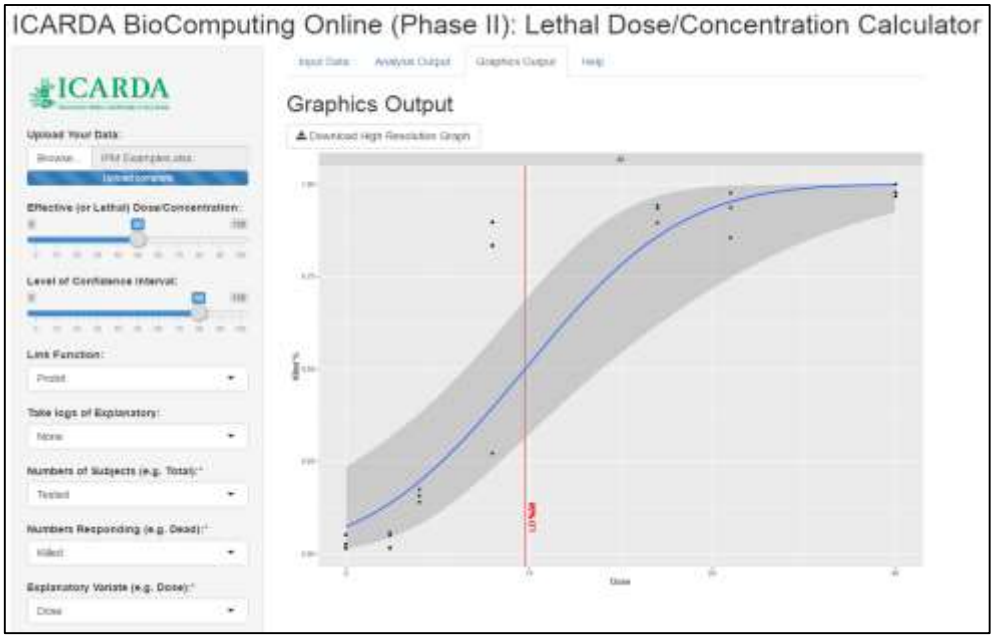

Fig.7. A screen shot from an online tool at ICARDA: LD50Shiny-3

References

Cochran, W.G. and Cox, G.M. 1957. Experimental Designs, (New York: John Wiley and Sons Inc.).

El-Shamaa, K. 2017. LD50 (Version 1.0) [Web application]. Biometrics and Statistics Section, ICARDA. http://geoagro.icarda.org/bss/shinyapps/ld50 (Accessed 12 Dec 2017).

Finney, D. J. 1952.Probit Analysis. Cambridge, England, Cambridge University Press

Gomez, K.A. and Gomez, A.A. 1984. *Statistical Procedures for Agricultural Research*. (New York: John Wiley and Sons Inc).

Hinkelmann, K.H., and Kempthorne, O. 2005.*Design and Analysis of Experiments*. Volume 2: Advanced Experimental Design. John Wiley & Sons, New York.

Singh, M. and El-Shamaa, K. 2015. Experimental Designs for Precision in Phenotyping. Pages 235-247, Chapter 16, in Phenomics in Crop Plants: Trends, Options and Limitations. Springer (Editors: Kumar, J., Pratap, A. and Kumar, S.) (doi: 10.1007/978-81-322-2226-2_16).

VSN International. 2015. The Guide to the Genstat Command Language (Release 18), Part 2 Statistics. VSN International, Hemel Hempstead, UK.

Zaid, A. and Botes, A. 2002. Chapter X: Establishment of a modern date plantation. *In* Date Palm Cultivation, FAO Plant Production and Protection Paper 156 Rev. 1. (Ed. Zaid,

A.). Food and Agricultural Organization of United Nations, Rome.http://www.fao.org/docrep/006/y4360e/y4360e0e.htm#bm14 (Accessed 19 Feb 2017).

Zaid, A., de Wet, P. F., Djerbi, M. and Oihabi, A. 2002. Chapter XII: Diseases and Pests of Date Palm. *In* Date Palm Cultivation, FAO Plant Production and Protection Paper 156 Rev. 1. (Ed. Zaid, A.). Food and Agricultural Organization of United Nations, Rome.

http://www.fao.org/docrep/006/y4360e/y4360e0g.htm#bm16.1 (Accessed 19 Feb 2017).