Strolling through the Bayesian modeling to assess plant disease epidemiology

Danilo Alvares[1], *Carmen Armero*[1], *Anabel Forte*[1], *and Luis Rubio*[2]

[1]daldasil@alumni.uv.es, carmen.armero@uv.es, anabel.forte@uv.es,
Department of Statistics and Operational Research, Universitat de València
[2]lrubio@ivia.es, Department of Virology, Valencian Institute for Agricultural Research

Abstract

The growing of tiger nuts (also known as *chufas*) is an important agricultural sector in València (Spain). Recently, this sector has been affected by the emergence of black spots on tubers' skin, which leads to the disposal of many of them. We explore some Bayesian models to analyze different relevant elements of the epidemiology of the disease.

Keywords: Dirichlet-multinomial model; Log-normal model; Measures of effectiveness.

1. Introduction

The presence of black spots in the skin of tiger nuts tubers is an emerging problem that affects an important agricultural sector in València. Tiger nuts tubers are the essential raw material for making the popular *horchata* [2], but tubers with black spots cannot be used for making this drink.

This work focuses on the statistical modeling to assess the relationships between the disease and the unit-weight of the harvested tubers, the transmission from infected seeds to harvested tubers, and the behavior and effectiveness of possible treatments for the disease.

2. Experiments and data

Data come from two different experiments carried out in greenhouse:

Experiment 1: Asymptomatic (group 1) and severely affected (group 2) tubers were separated and sowed in different pots with five seeds per pot. In particular, the experiment was replicated 7 times for each group. Five months after sowing, tubers were harvested and 100 tubers from each pot were randomly sampled. From them the mean unit-weight, in grams, and the number of tubers in each level of affection (asymptomatic, mild symptoms and severe symptoms) were measured.

Experiment 2: Chemical and/or thermal treatments were applied to severely affected seeds following a balanced two-factor factorial design. The specific combination of these treatments is denoted by $\mathbf{T}_{qt}$, where sub-index q represents chemical treatments: No treatment ($q = 1$), Sodium hypochlorite ($q = 2$), Hydrochloric acid ($q = 3$), Trifloxystrobin (fungicide) ($q = 4$), Acibenzolar-S-methyl (BION) ($q = 5$), and Trisodium phosphate ($q = 6$); and sub-index t is for thermal treatments: No treatment ($t = 1$), 55^{o}C for 30 min ($t = 2$), and 60^{o}C for 30 min ($t = 3$). We sowed eight pots with five seeds each for every $\mathbf{T}_{qt}$ treatment. About five months later tubers were harvested and the number of tubers in each level of affection (asymptomatic, mild symptoms and severe symptoms) were measured. No data was collected for $\mathbf{T}_{33}$ since no seed germinated.

3. Models and results

Different Bayesian models were used to analyze data from both experiments according to the nature and objectives of the study.

3.1. Models for positive continuous response variables

Our first focus is the variable measuring the mean unit-weight of tubers from *Experiment 1*. We compare several popular models for positive continuous response variables implemented with their usual non-informative priors [3]. Namely, we explore log-normal, gamma, exponential, chi-squared, inverse gamma, and Weibull models. Deviance information criterion (DIC) chooses the log-normal model as the best model (among considered) to explain data. Figure 1 shows the subsequent posterior distribution from the log-normal model in each experimental group.

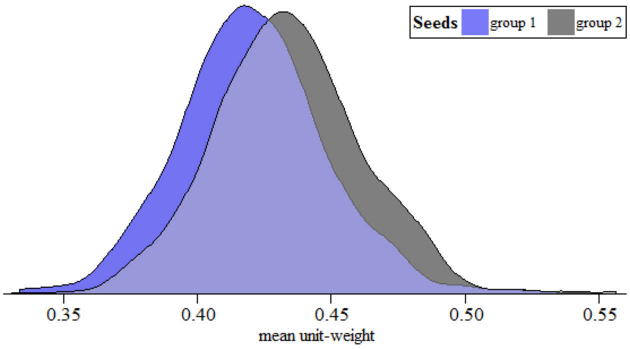

Figure 1: Posterior distribution for the mean unit-weight, in grams, of the tubers harvested from asymptomatic seeds (group 1) and seeds with severe symptoms black spots (group 2).

Results indicate that both posterior distributions are almost identical, with 95% credible intervals $(0.368,\ 0.477)$ and $(0.381,\ 0.488)$ grams in groups 1 and 2, respectively.

3.2. Dirichlet-multinomial model

We continue to work with *Experiment 1* and study the number of harvested tubers in each level of affection (asymptomatic, with mild and severe symptoms of the disease) for each group. We use a Dirichlet-multinomial model [1] and the Perks Dirichlet objective prior [4] for the probabilities of the multinomial distribution. Figure 2 shows a 95% credible interval for the marginal posterior distribution of the relevant probabilities in each experimental group.

Results in group 1 indicate that the probability of harvesting asymptomatic tubers (posterior mean 0.6976) is greater than the probabilities corresponding to tubers with mild (posterior mean 0.2283) and severe symptoms (posterior mean 0.0741). The situation in group 2 is the opposite, with a probability of harvesting tubers with severe symptoms (posterior mean 0.8030) larger than the probabilities corresponding to tubers with mild symptoms (posterior mean 0.1968) and asymptomatic (posterior mean 0.0002).

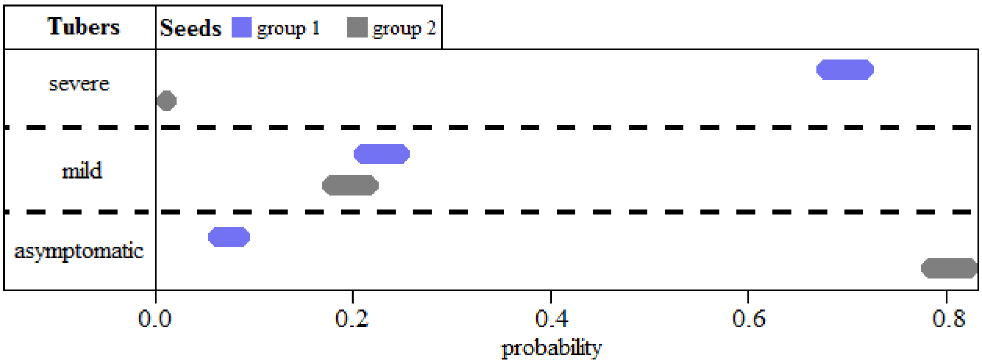

Figure 2: 95% credible interval for the probabilities associated to tubers harvested from asymptomatic seeds (group 1), and seeds with severe symptoms black spots (group 2).

3.3. Logistic regression models and measures of effectiveness

We analyze the effectiveness of the different chemical and thermal treatments of *Experiment 2* for reducing black spots in harvested tubers when using black spot tubers as seeds. Bayesian logistic regression models [1] are used for assessing the harvested number of asymptomatic tubers and the number of germinated seeds for each combination, $\mathbf{T}_{qt}$, of chemical and thermal treatment. Treatments $\mathbf{T}_{23}$ and $\mathbf{T}_{53}$, both based on a temperature of $60^o C$, achieve the best results with regard to the probability of obtaining asymptomatic tubers. Treatment $\mathbf{T}_{31}$ and, to a lesser extent, $\mathbf{T}_{11}$, $\mathbf{T}_{21}$, and $\mathbf{T}_{41}$ have the greatest values of the probability of seed germination and none of them include thermal treatment.

The results for both quantities of interest appear to be antagonistic. To find the best on balance treatment we construct a measure of effectiveness

$$\theta_{eqt} = v\,\theta_{1qt} + (1-v)\,\theta_{2qt},$$

that weights the probability, θ_{1qt}, of harvesting asymptomatic tubers and the probability of seed germination, θ_{2qt}, with regard to the particular interest of the farmer with $0 \leq v \leq 1$ as the weighting constant in favor of the probability of harvesting asymptomatic seeds.

Table 1 shows the posterior mean of θ_{eqt} for the different treatments and some elicited values of v. When the priority is the germination, the most effective treatment is $\mathbf{T}_{31}$ (chemical treatment with hydrochloric acid and without thermal treatment); but if the priority is the number of asymptomatic tubers, the best option is $\mathbf{T}_{21}$ (chemical treatment with sodium hypochlorite and without thermal treatment). In the case of a balanced decision ($v = 0.5$), the best and worst options are treatments $\mathbf{T}_{21}$ and $\mathbf{T}_{53}$, respectively.

4. Conclusions

We have applied different Bayesian models for discussing some epidemiological elements of an unknown disease affecting tiger nuts. Results indicated that the mean unit-weight of har-

Table 1: Posterior mean of the measure of effectiveness θ_{eqt} for some values of the weighting constant v.

v	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
$1-v$	**0.8**	**0.7**	**0.6**	**0.5**	**0.4**	**0.3**	**0.2**
$\mathbf{T}_{11}$	0.370	0.344	0.318	0.292	0.266	0.239	0.213
$\mathbf{T}_{12}$	0.309	0.327	0.344	0.362	0.379	0.397	0.414
$\mathbf{T}_{13}$	0.116	0.161	0.206	0.251	0.296	0.341	0.386
$\mathbf{T}_{21}$	0.455	0.460	0.465	0.470	0.475	0.480	0.485
$\mathbf{T}_{22}$	0.300	0.314	0.328	0.341	0.355	0.369	0.382
$\mathbf{T}_{23}$	0.176	0.226	0.277	0.327	0.378	0.428	0.479
$\mathbf{T}_{31}$	0.549	0.509	0.470	0.431	0.392	0.352	0.313
$\mathbf{T}_{32}$	0.205	0.194	0.183	0.172	0.161	0.151	0.140
$\mathbf{T}_{41}$	0.430	0.406	0.383	0.360	0.337	0.314	0.291
$\mathbf{T}_{42}$	0.254	0.245	0.236	0.227	0.218	0.209	0.201
$\mathbf{T}_{43}$	0.156	0.208	0.261	0.313	0.366	0.418	0.471
$\mathbf{T}_{51}$	0.374	0.375	0.375	0.376	0.377	0.378	0.378
$\mathbf{T}_{52}$	0.362	0.357	0.352	0.348	0.343	0.338	0.333
$\mathbf{T}_{53}$	0.059	0.075	0.092	0.108	0.125	0.141	0.158
$\mathbf{T}_{61}$	0.317	0.312	0.308	0.304	0.299	0.295	0.291
$\mathbf{T}_{62}$	0.188	0.206	0.225	0.243	0.262	0.280	0.299
$\mathbf{T}_{63}$	0.087	0.105	0.123	0.142	0.160	0.178	0.196

vested tubers is not affected by the fact that the seeds are asymptomatic or not; conversely, the state of health of the seeds seriously affects the health of the harvested tubers. When using some treatment in advance to sowing seeds we conclude that when prioritizing the germination the treatment with hydrochloric acid is the most appropriate; on the other hand, when the priority is the rate of asymptomatic tubers harvested the best option is using a treatment with sodium hypochlorite.

5. Acknowledgments

Alvares's work was supported by Coordination for the Improvement of Higher Level Personnel (BEX: 0047/13–9), Brazil. Armero and Forte's work was supported by Grant MTM2013–42323–P from the Spanish Ministry of Economy and Competitiveness. Rubio's work was supported by Valencian Institute for Agricultural Research, Spain.

6. Bibliography

[1] Congdon, P. (2005). *Bayesian models for categorical data.* Probability and Statistics. 1st ed. John Wiley and Sons.

[2] Morell, J. and Barber, S. (1983). Chufa y horchata: características físicas, químicas y nutritivas. *Technical report.* Institute of Agrochemistry and Food Technology, Valencia, Spain.

[3] Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS.* Computational Statistics. John Wiley and Sons. Hoboken, USA.

[4] Perks, W. (1947). Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries*, 73:285–334.