Volumen 31, Nº 2. Páginas 107-110
IDESIA (Chile) Mayo-Agosto, 2013

Survival analysis in plant pathology

Análisis de supervivencia en fitopatología

Cristiano Nunes Nesi[1], Silvia Emiko Shimakura[2], Paulo Justiniano Ribeiro Junior[2], Louise Larissa May De Mio[3]*

ABSTRACT

Survival analysis is applied when the time until the occurrence of an event is the object of interest. Such data are routinely collected in plant pathology whereas application of methods of survival analysis is uncommon. Basic concepts in survival analysis for use in plant pathology such as Kaplan-Meier methods, log-rank test and Cox proportional hazards models are introduced and applied estimating the effects of cultivars on the survival times of brown rot symptoms and on the instantaneous risk of expressing symptoms in a hypothetical study.

Key words: time-occurrence, Kaplan-Meier, Cox regression.

RESUMEN

El análisis de supervivencia se aplica cuando el tiempo hasta la ocurrencia de un evento es el objeto de interés. En enfermedades vegetales, tales datos se recogen habitualmente, aunque aplicación no es común. El objetivo de este trabajo fue introducir conceptos básicos de análisis de supervivencia para su uso en patología vegetal. Fueron utilizados métodos de Kaplan-Meier, prueba log-rank y riesgos proporcionales de Cox para estimar el efecto de cultivares en la expresión de los síntomas de la pudrición parda y sobre el riesgo instantáneo de expresar en un estudio hipotético.

Palabras clave: *tiempo-ocurrencia, Kaplan-Meier, regresión de Cox.*

The most common mathematical tools used in plant disease epidemiology are disease progress curves, linked differential equations, area under disease progress curve and computer simulations (Madden *et al.*, 2007; Contreras-Medina *et al.*, 2009). In plant pathology, the time to occurrence of events such as infections is routinely collected on laboratories or field trials. However, adoption of suitable methods of statistical survival analysis is unusual (Garrett *et al.*, 2004; Scherm and Ojiambo, 2004). Most commonly, data on a single date of assessment are used to compare treatments using analysis of variance, possibly after transformation of the response variable given from time to event. Such popular statistical technique become inappropriate because the time to event rarely follows a normal distribution and frequently the study ends before all individuals suffer the event, leading to censored responses (Rebasa, 2005). Both, removal or naïve usage of the censored data as observed time introduces bias, limits the ability of correct inference and reduces the power of statistical tests. In this context survival analysis is appropriate to build estimators for the functions describing the survival time and allowing for testing the influence of experimental factors (Carvalho *et al.*, 2011).

Examples of survival analysis applied to plant pathology are rarely observed in the literature. Scherm and Ojiambo (2004) reviewed important principles and procedures of survival analysis and modeled the time to abscission of blueberry leaves in relation to Septoria leaf spot severity, leaf age and leaf location in the canopy in order to provide

[1*] Corresponding author; Federal University of Paraná/UFPR; Department of Crop Protection, Rua dos Funcionários, 1540, 80.035-050, Curitiba-PR-Brazil.
[2] Federal University of Paraná-UFPR; Department of Statistics, Centro Politécnico, 81.531-990, Curitiba/PR, Brazil.
[3] Federal University of Paraná-UFPR; Department of Crop Protection, Rua dos Funcionários, 1540, 80.035-050, Curitiba/PR, Brazil.
* Corresponding author: cristiano@epayi.ac.gov.br

Fecha de Recepción: 25 Enero, 2013.
Fecha de Aceptación: 02 Abril, 2013.

estimates of the probability of leaf loss at given times during the growing season. Survival analysis allowed to explore and quantify the effects of tree location, peach orchard attributes and disease status in the vicinity of diseased orchards upon the risk for the trees to become infected by *Plum pox virus* strain M through time and, as a result, affecting the persistence of the disease (Dallot *et al.*, 2004).

Gottwald and Taylor (2005) examine the probability of trees remaining at non-infected state by *Citrus tristeza virus* (CTV) considering various distances of proximity to CTV-infected trees. Using nonparametric and parametric methods of survival analysis, Esker *et al.* (2006) tested different hypotheses regarding plant factors that may influence the post incubation survival time of phytoplasma-infected papaya.

Survival analysis using Kaplan-Meier estimates, Cox proportional hazards and extended Cox models were used to determine potential seasonal differences in incubation periods of Camellia twig blight, caused by *Colletotrichum gloeosporioides*. The time until appearance of the first symptom of disease (incubation period length) was recorded and stems without disease symptoms by the last day of the observation period were recorded as censored observations (Copes and Thomson, 2008). Setti *et al.* (2010) investigate the use of survival methods to estimate the latent period of *Mycospherella pinodes* infected peas and to assess the effect of isolates aggressiveness, leaf wetness duration, inoculum concentration, plant age and host susceptibility on time to pycnidia formation.

To illustrate the use of survival analysis in plant pathology, we consider a study comparing the time to express brown rot symptoms caused by *Monilinia fructicola* of post-harvest fruits of three peach cultivars. Data analyses were performed using the R statistical system (R Development Core Team, 2011) and the add-on package 'survival' (Therneau, 2012).

The main non-parametric technique in survival analysis is the Kaplan-Meier estimator (Kaplan and Meier, 1958), computed by updating the survival function after each event, without the assumption of an underlying probability distribution (Bewick *et al.*, 2004; Rebasa, 2005). Figure 1a shows a plot of an estimated survival function against time called the survival curve and represented by steps to indicate the instants of time in which events occur and signs (+) to indicate censoring. Influence of experimental factors upon survival times can be assessed by fitting a semi-parametric regression model known as Cox model and random effects expressing individual frailty can be included (McGilchrist and Aisbett, 1991). This model is a proportional hazards model because the hazard rates for individuals with different covariate values are assumed constant over time. The assumption of independence among times of occurrence is invalid if unobserved characteristics are shared by groups of experimental units. In such cases a frailty model captures the within-group dependence by assuming a shared unobservable random effect, resulting in better estimation of the effects of experimental factors.

For our example, to test the hypothesis that survival time was affected by different cultivars we have applied the Kaplan-Meier estimator of survival probabilities, using the log-rank test. In this case, the estimated survival functions were significantly different ($p = 8.55\ 10^{-8}$) between cultivars (Figure 1a). A Cox model fitted to these data suggests that cultivars B and C are, respectively, 2.7 and 5.0 times more likely to express symptoms at any time than cultivar A. The analysis supports the proportional hazards assumption by inspection of the Kaplan-Meier curves (Figure 1a) and Schoenfeld residual plots (Figures 1c and 1d). The correlations between these residuals and the standardized time are non-significant. The estimated variance of the random effects –frailty– is non-significant (Figure 1b) suggesting that expression of symptoms is independent among fruits from the same tree.

Survival methods can enhance the statistical analysis of experiments where time to events are recorded by taking better account of the data structure and allowing for richer and more precise inference. Further models of survival data, allowing for different assumptions, with the occurrence of the event prior to observation in the individual (left censoring) or within an interval time (interval censoring) can also be modeled. When the proportional hazards assumption in the Cox model is not fulfilled alternative models can be used such as accelerated failure time models. Furthermore, the relationship between continuous covariates and the associated risks can be estimated with nonparametric functions.

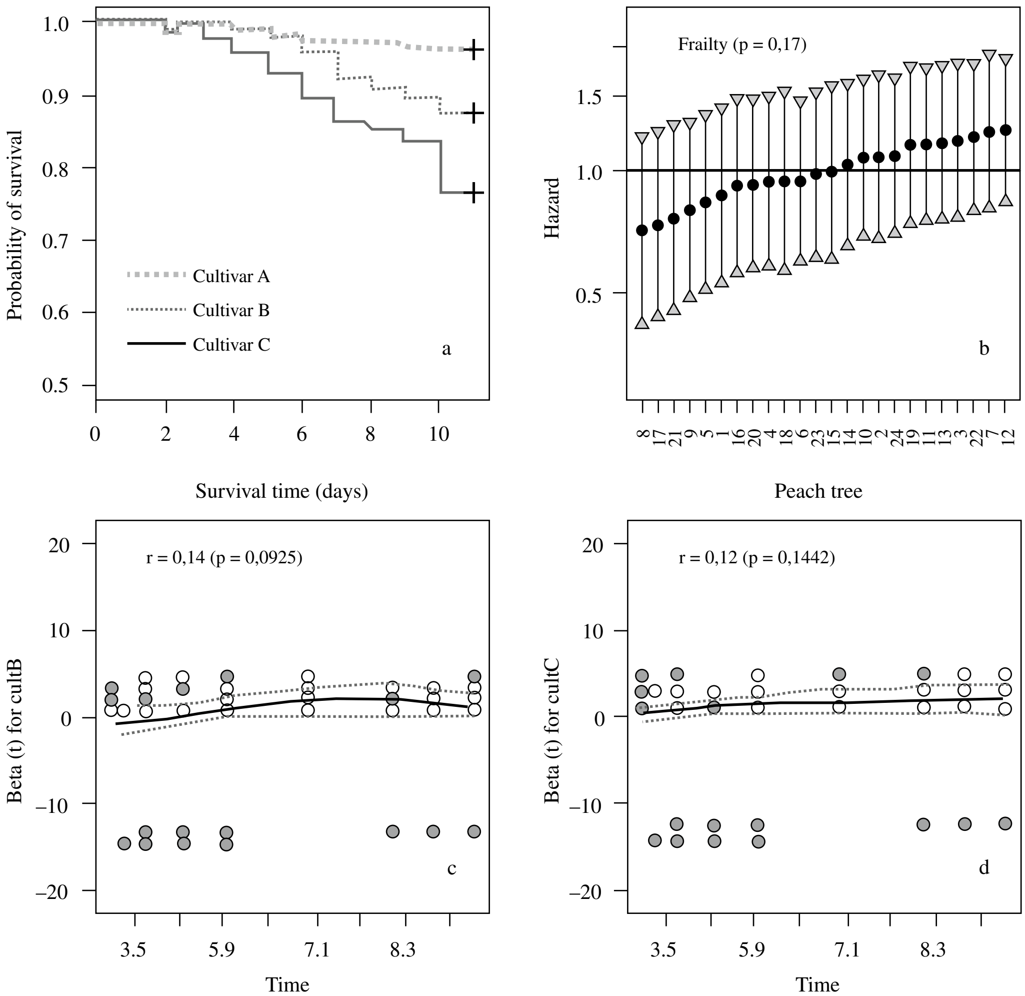

Figure 1. The Kaplan-Meier survival curves (a), estimates of frailty and their 95% confidence intervals (b), Schoenfeld residual plots (c and d) and correlations between these residuals and the standardized time for the cultivars B and C. The analysis refers to a study which evaluated the time (days) for expression brown rot symptoms on fruit of post-harvest in peach cultivars.

References

Bewick, V.; Cheek, L.; Ball, J.
2004. Statistics review 12: Survival analysis. *Critical Care* 8 (5): 389-394.

Carvalho, M.S.; Andreozzi, V.L.; Codeço, C.T.; Campos, D.P.; Barbosa, M.T.S.; Shimakura, S.E.
2011. Análise de Sobrevivência: teoria e aplicações em saúde, 2a. Edición. Rio de Janeiro. FIOCRUZ, 432 p.

Contreras-Medina, L.M.; Torres-Pacheco, I.; Guevara-González, R.G.; Romero-Troncoso, R.J.; Terol-Villalobos, I.R.; Osornio-Rios, R.A.
2009. Mathematical modeling tendencies in plant pathology. *African Journal of Biotechnology* 8 (25): 7399-7408.

Copes, W.E.; Thomson, J.L.
2008. Survival analysis to determine the length of the incubation period of Camellia twig blight caused by *Colletotrichum gloeosporioides*. *Plant Disease* 92: 1177-1182.

Dallot, S. Gottwald, T. Labonne, G. Quiot, J.B.
2004. Factors affecting the spread of *Plum pox virus* strain M in peach orchards subjected to roguing in France. *Phytopathology* 94: 1390-1398.

Esker, P.D.; Gibb, K.S.; Padovan, A.; Dixon, P.M.; Nutter Jr., F.W.
2006. Use of survival analysis to determine the postincubation time-to-death of papaya due to yellow crinkle disease in Australia. *Plant Disease* 90: 102-107.

Garrett, K.A.; Madden, L.V.; Hughes, G.; Pfender, W.F.
2004. New applications of statistical tools in plant pathology. *Phytopathology* 94: 999-1003.

Gottwald, T.R.; Taylor, E.L.
2005. Using survival analysis to predict the risk of infection in a Citrus tristeza virus epidemic. In: Proc. Sixteenth IOCV Conference 101-111.

Kaplan, E.L.; Meier, P.
1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53 (282): 457-481.

Madden, L.V.; Hughes, G.; Bosh, F.V.D.
2007. The study of plant disease epidemics. Minnesota: The American Phytopathological Society.

McGilchrist, C.A.; Aisbett, C.W.
1991. Regression with frailty in survival analysis. *Biometrics* 47: 461-466.

Rebasa, P.
2005. Conceptos básicos del análisis de supervivencia, *Cirurgía Española*, 78 (4): 222-230.

Scherm, H.; Ojiambo, P.S.
2004. Applications of survival analysis in botanical epidemiology, *Phytopathology*, 94: 1022-1026.

R Development Core Team.
2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [http://www.R-project.org/]. Accessed 10 October 2011.

Setti, B.; Bencheikh, M.; Henni, J.E.; Claire, N.
2010. Survival analysis to determine the length of latent period of *Mycosphaerella pinodes* on peas (*Pisum sativum* L.). *African Journal of Microbiology Research* 4 (18): 1897-1903.

Therneau, T.
2012. A Package for Survival Analysis in S. R package version 2. 36-12.