29 March 2019
Rgronomists/LunchinatoRs

Survival Analysis in R

Vivian Bernau, PhD

Maize Curator, USDA-ARS Plant Introduction Station

vivian.bernau@usda.gov

THE OHIO STATE UNIVERSITY

COLLEGE OF FOOD, AGRICULTURAL, AND ENVIRONMENTAL SCIENCES

USDA

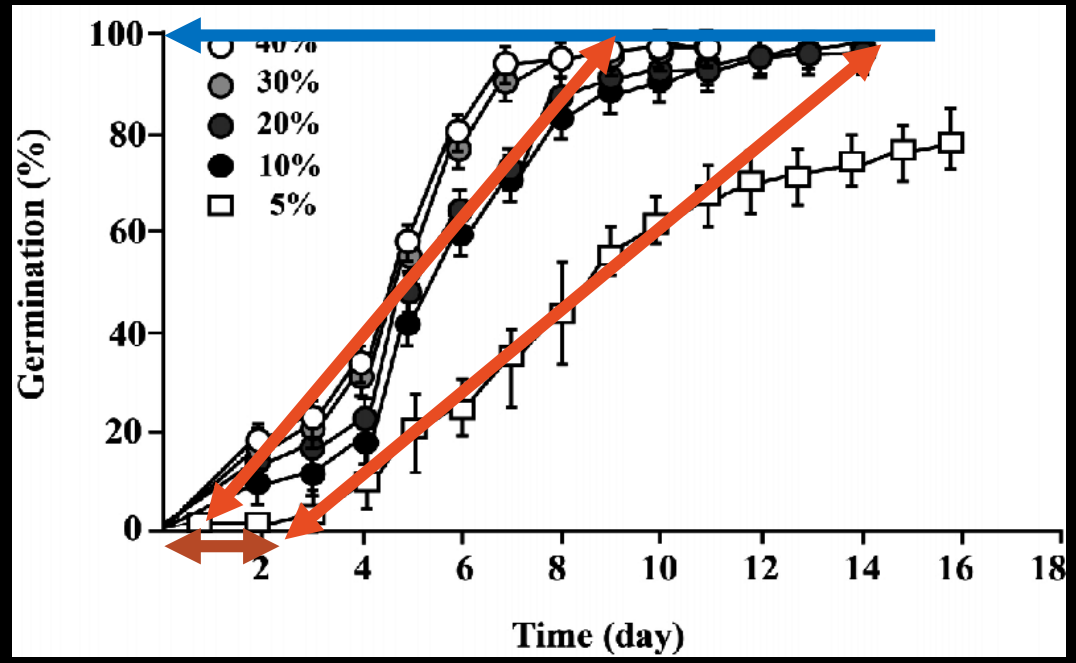
ARS

What is survival analysis?

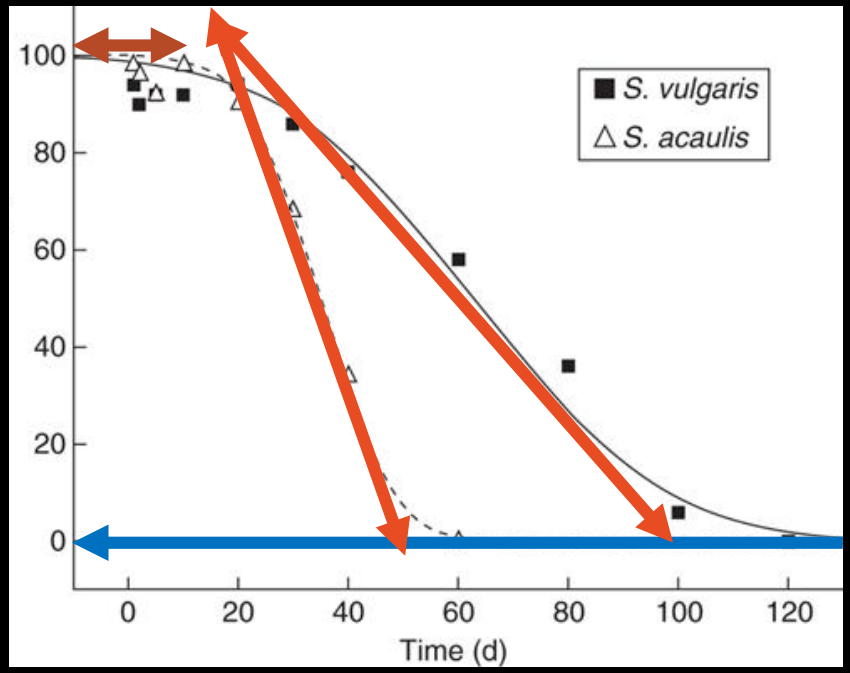
- Also known as time-to-event analysis
- Analysis of the expected duration of time until one or more events happen
  - death of an organism
  - failure of a machine component
  - *germination of a seed*
- 1 subject = 1 event
- Data is ”right-censored”

A survival curve is the inverse of a germination curve.

Germination Curve



Survival Curve

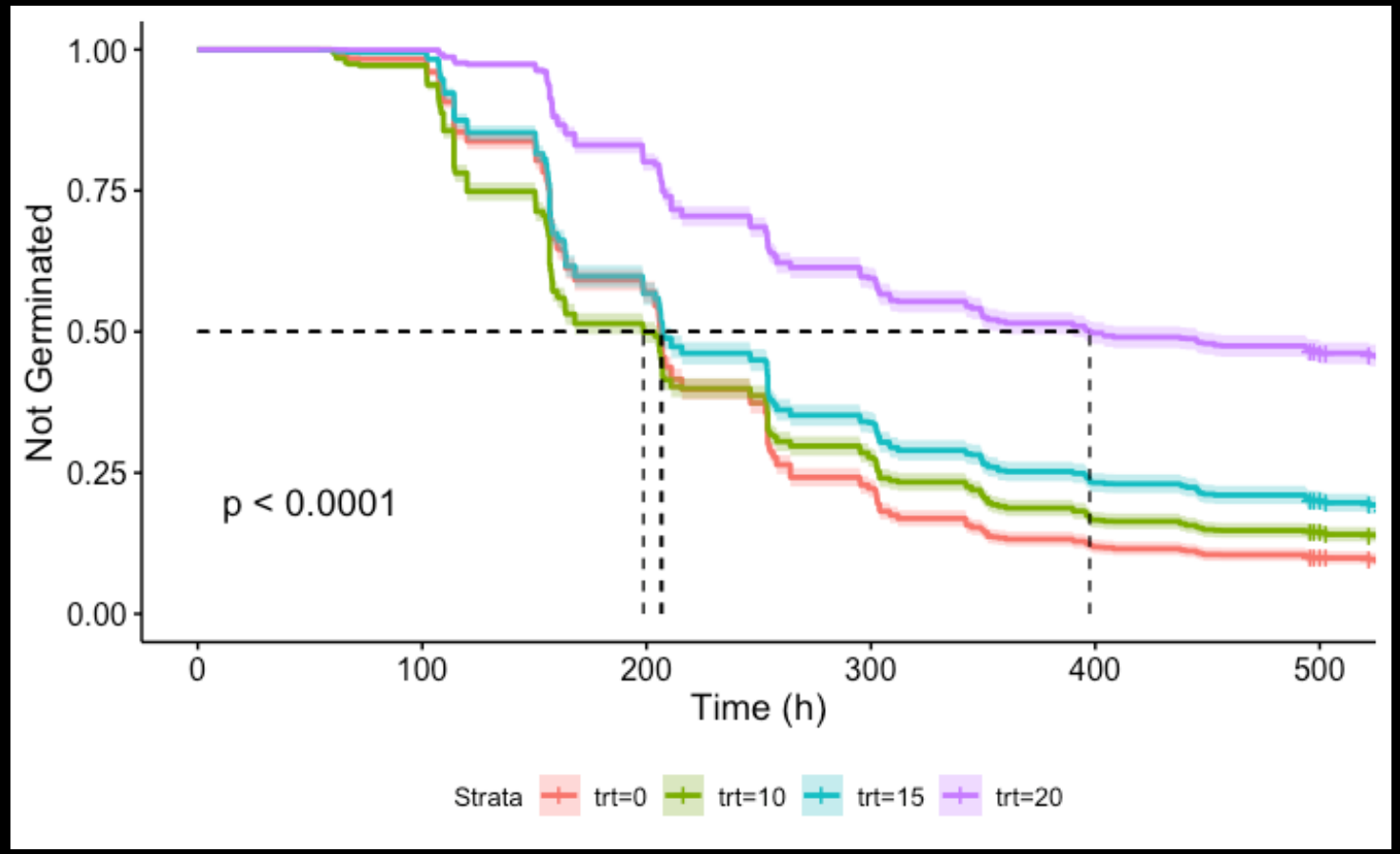


100% germination

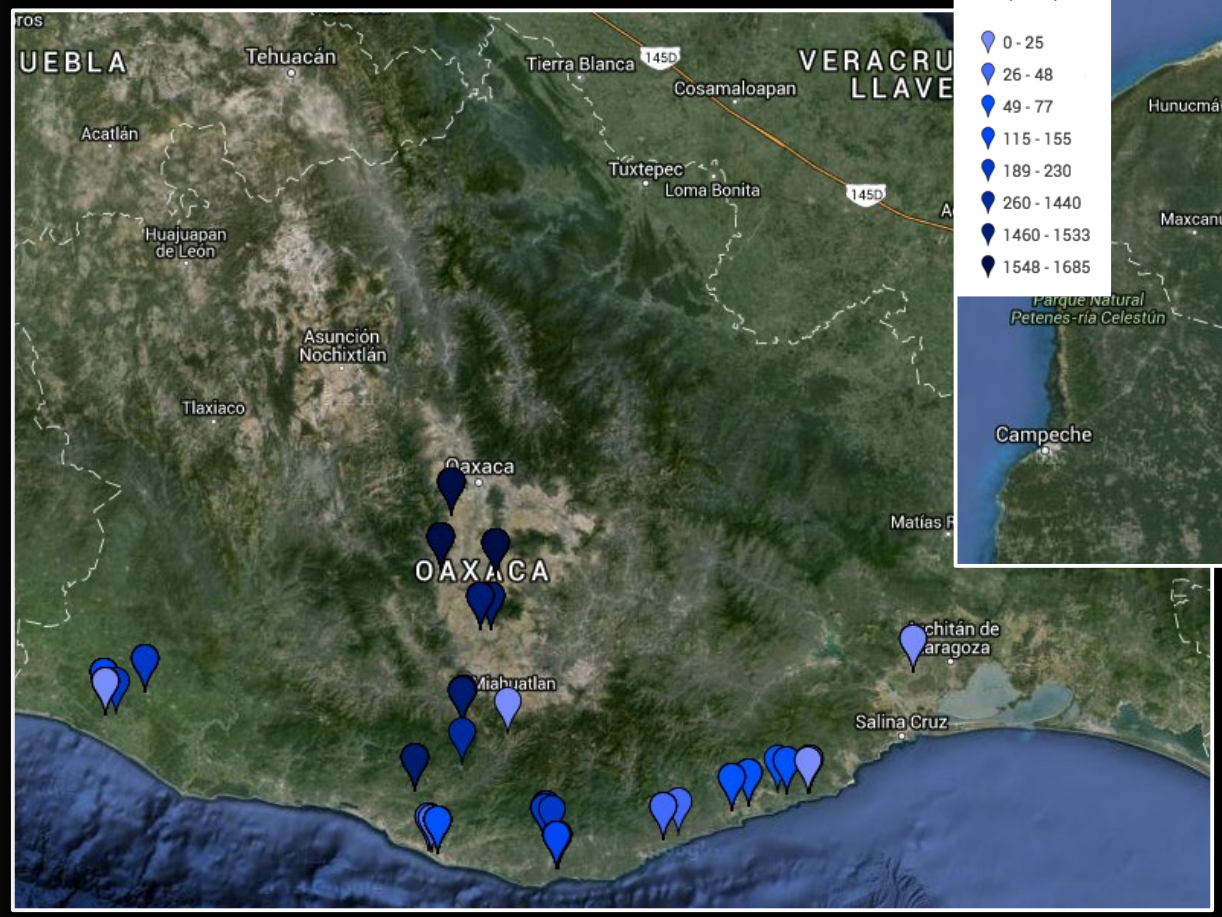
Germination rate

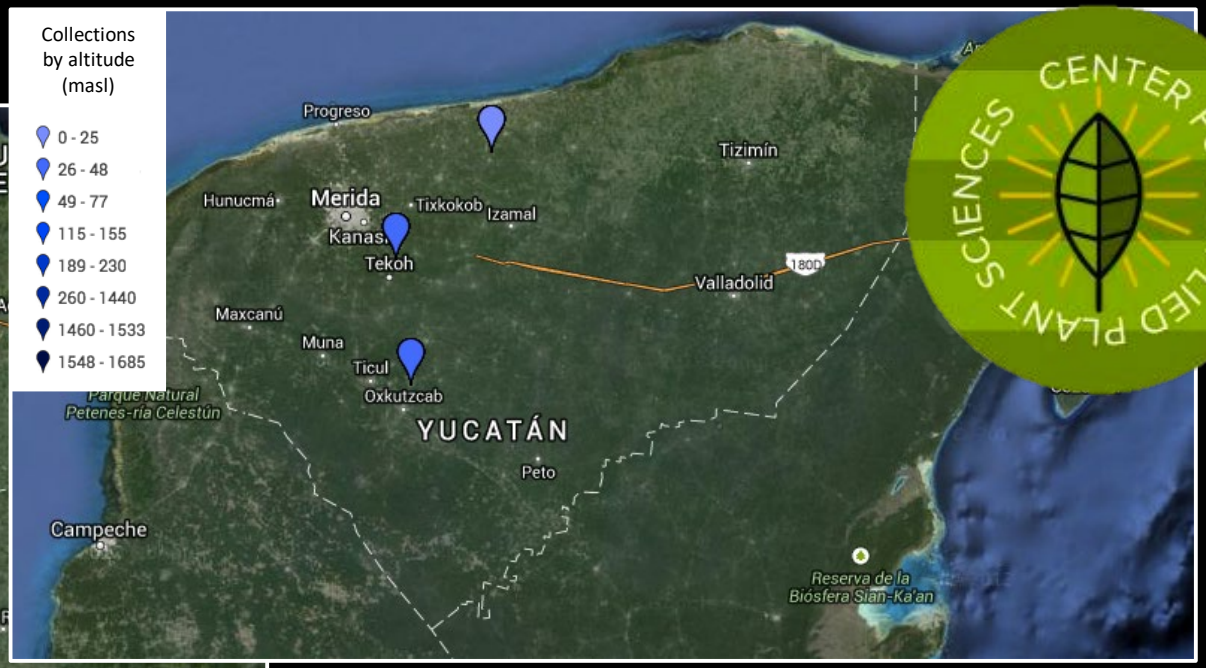
Delay

Kaplan-Meier Curve



- *S(t)* = probability that subject survives longer than time *(t)*.
- *S(t)* is estimated with the Kaplan Meier curve. (step-like estimate)





CULTIVATED **WILD**

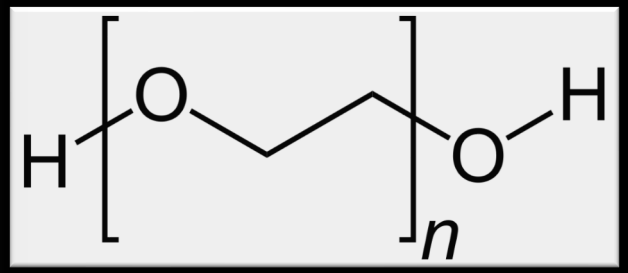
***CAPS* Chile Collections 2013-2014**

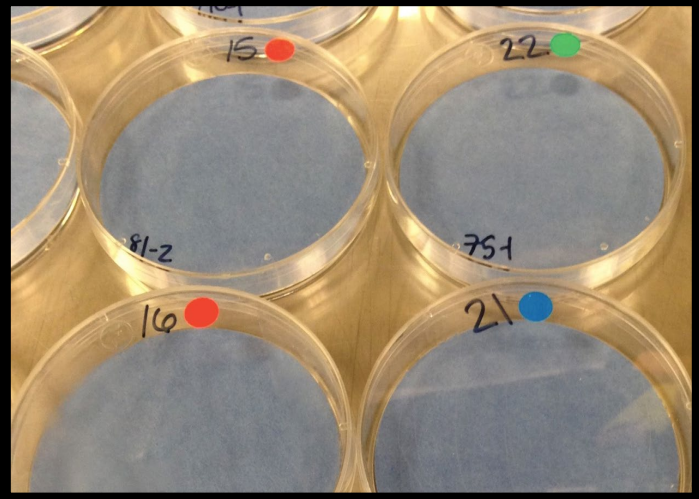
Collection gradients: geographic, elevational, climatic, ethnic, domestication

Objective: Assess germination response of diverse chiles under osmotic stress

- Evaluated 131 lines (18 landraces, 36 populations)
- 4 Solutions: 0% PEG, 10%, 15%, and 20% PEG

Polyethylene Glycol





Packages: ‘survival’ & ‘survminer’

- ‘survival’
  - was developed in S by Terry Therneau, a clinical research statistician at the Mayo Clinic
  - contains core survival analysis routines: definition of survival objects, **Kaplan-Meier** and Aalen-Johansen curves, Cox models, and parametric accelerated failure time models.
- ‘survminer’
  - Developed by Alboukadel Kassambara, as cancer bioinformatics scientist
  - contains ‘ggsurvplot()’ which creates attractive survival curves

My code

Run the survival function.

```
#fit and plot Kaplan-Meier survivor function, includes 95% confidence intervals
test.peg <- survfit(Surv(end, status) ~ trt ,data = df, type = "kaplan-meier")
ggsurvplot(test.peg, data =df, conf.int = T, pval = T,
           risk.table = F, xlab = "Time (h)", ylab = "Not Germinated",
           surv.median.line = "hv", legend = "bottom")
ggsave("peg.jpg", width = 6, height = 4, units = "in", dpi = 300)
survdiff.peg <- pairwise_survdiff(Surv(end, status)~trt, data = df, p.adjust.method = "bonferroni",rho = 0)
sink("peg.txt")
print(survdiff.peg)
sink()
```

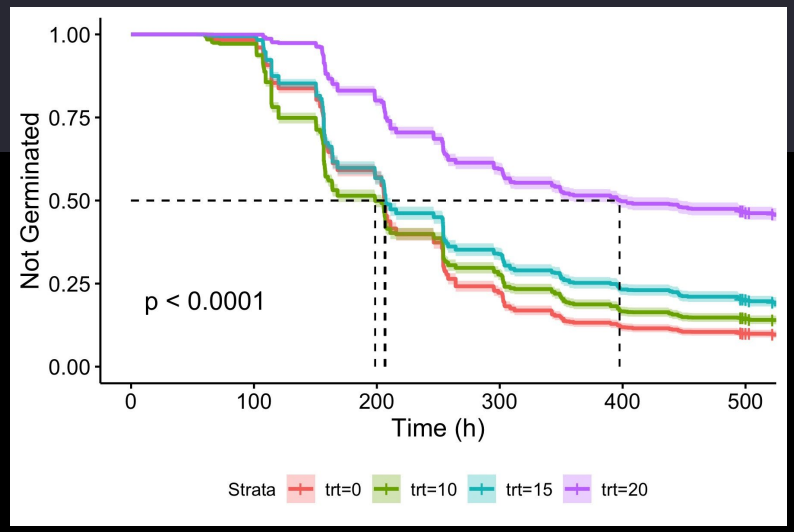
My code

Plot the survival function.

```r
#fit and plot Kaplan-Meier survivor function, includes 95% confidence intervals
test.peg <- survfit(Surv(end, status) ~ trt ,data = df, type = "kaplan-meier")
ggsurvplot(test.peg, data =df, conf.int = T, pval = T,
           risk.table = F, xlab = "Time (h)", ylab = "Not Germinated",
           surv.median.line = "hv", legend = "bottom")
ggsave("peg.jpg", width = 6, height = 4, units = "in", dpi = 300)
survdiff.peg <- pairwise_survdiff(Surv(end, status)~trt, data = df, p.adjust.method = "bonferroni",rho = 0)
sink("peg.txt")
print(survdiff.peg)
sink()
```

My code

```
#fit and plot Kaplan-Meier survivor function, includes 95% confidence intervals
test.peg <- survfit(Surv(end, status) ~ trt ,data = df, type = "kaplan-meier")
ggsurvplot(test.peg, data =df, conf.int = T, pval = T,
           risk.table = F, xlab = "Time (h)", ylab = "Not Germinated",
           surv.median.line = "hv", legend = "bottom")
ggsave("peg.jpg", width = 6, height = 4, units = "in", dpi = 300)
survdiff.peg <- pairwise_survdiff(Surv(end, status)~trt, data = df, p.adjust.method = "bonferroni",rho = 0)
sink("peg.txt")
print(survdiff.peg)
sink()
```



Save the image in your out directory
at preferred size and resolution.

My code

```r
#fit and plot Kaplan-Meier survivor function, includes 95% confidence intervals
test.peg <- survfit(Surv(end, status) ~ trt ,data = df, type = "kaplan-meier")
ggsurvplot(test.peg, data =df, conf.int = T, pval = T,
           risk.table = F, xlab = "Time (h)", ylab = "Not Germinated",
           surv.median.line = "hv", legend = "bottom")
ggsave("peg.jpg", width = 6, height = 4, units = "in", dpi = 300)
survdiff.peg <- pairwise_survdiff(Surv(end, status)~trt, data = df, p.adjust.method = "bonferroni",rho = 0)
sink("peg.txt")
print(survdiff.peg)
sink()
```

Compare Kaplan-Meier estimators pairwise by treatment, use the Bonferroni method to adjust p.

My code

Capture output of pairwise comparisons as a text file.

```
#fit and plot Kaplan-Meier survivor function, includes 95% confidence intervals
test.peg <- survfit(Surv(end, status) ~ trt ,data = df, type = "kaplan-meier")
ggsurvplot(test.peg, data =df, conf.int = T, pval = T,
           risk.table = F, xlab = "Time (h)", ylab = "Not Germinated",
           surv.median.line = "hv", legend = "bottom")
ggsave("peg.jpg", width = 6, height = 4, units = "in", dpi = 300)
survdiff.peg <- pairwise_survdiff(Surv(end, status)~trt, data = df, p.adjust.method = "bonferroni",rho = 0)
sink("peg.txt")
print(survdiff.peg)
sink()
```

```
        Pairwise comparisons using Log-Rank test

data:  df and trt

   0        10       15
10 0.48     -        -
15 < 2e-16 4.2e-11 -
20 < 2e-16 < 2e-16 < 2e-16

P value adjustment method: bonferroni
```

National Plant Germplasm System mission:

CHARACTERIZE

CONSERVE

ACQUIRE

DOCUMENT

DISTRIBUTE

David Gilliver, *Harvesting the Corn*

CROP GERMPLASM

CIMMYT Station at Metepec

Maize seedling emerging from 20cm planting depth after 14 days; 2600masl

Field Site near Nevado de Toluca; 3400masl

Vivian Bernau

vivian.bernau@usda.gov

@vbern