SURVIVAL CURVES

Dr. Aric LaBarr

Institute for Advanced Analytics

MSA Class of 2014

INTRODUCTION

What is Survival Analysis?

- How long before something happens?
  - Study of occurrence and timing of events.
- Event is a qualitative change that can be tied to a specific point in time.
- Originally designed to study the occurrence of death in medical studies – hence *survival analysis*.
- Other names:
  - Time-to-event analysis
  - Duration analysis
  - Failure time analysis

Modern Examples

- Survival analysis is perfect for answering business questions, especially involving customers.
- When to start worrying about customers doing something important?
  - When will a customer cancel service?
  - When will a customer make another purchase?
- What factors are correlated with different events?
  - How many customers will we lose is we raise prices by 10%?
  - What impact can we expect from a loyalty program?

SURVIVAL CURVES

Survival Curves

- Proportion of observations that are expected to “survive” up to a particular point in time.
  - Always starts at 1 (or 100%).
  - Never increases.
  - Rarely goes to zero.
  - After an observation has stopped, it cannot come back.

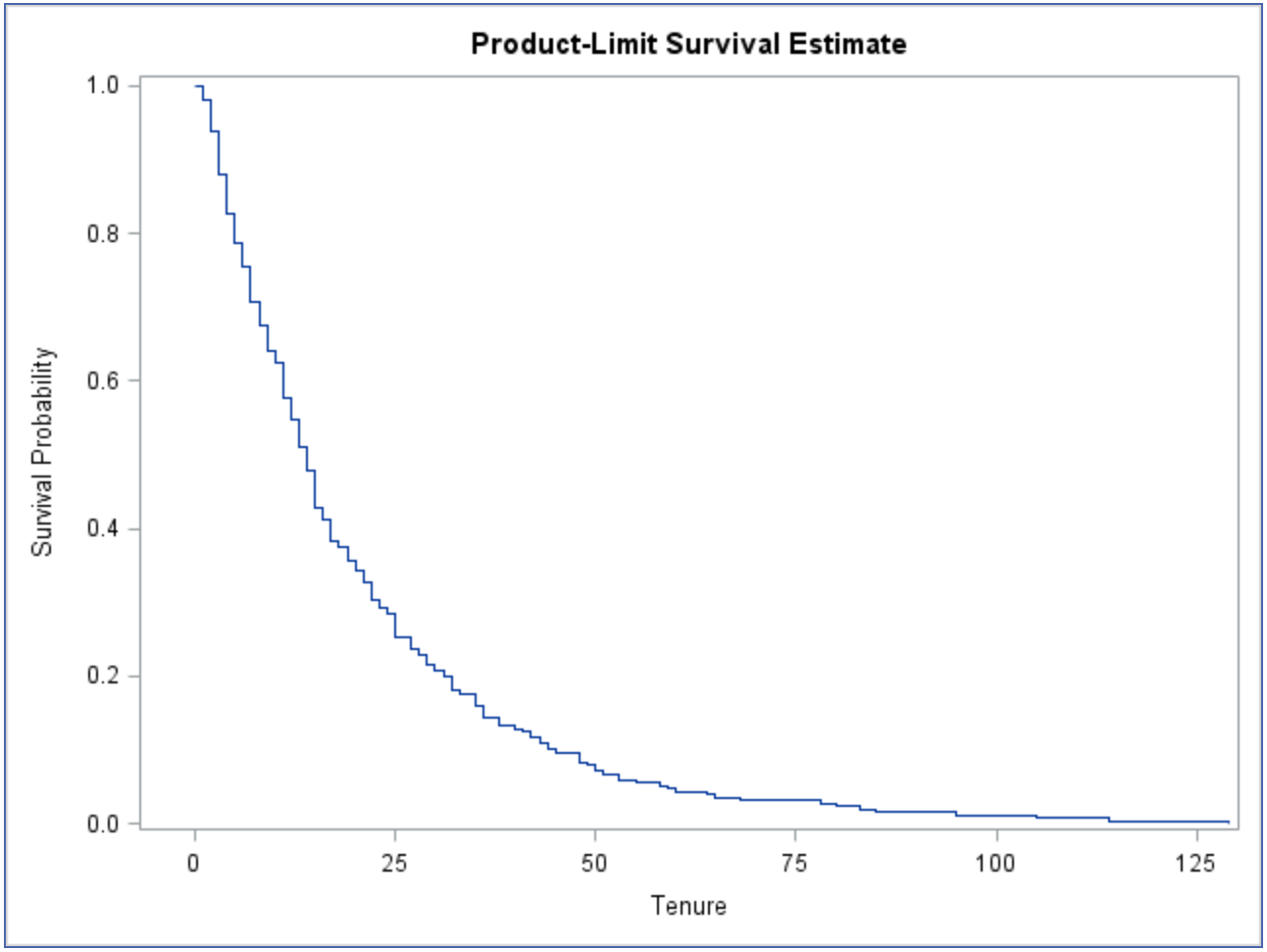
- Survival curves use to be the only method in survival analysis.

Survival Curves

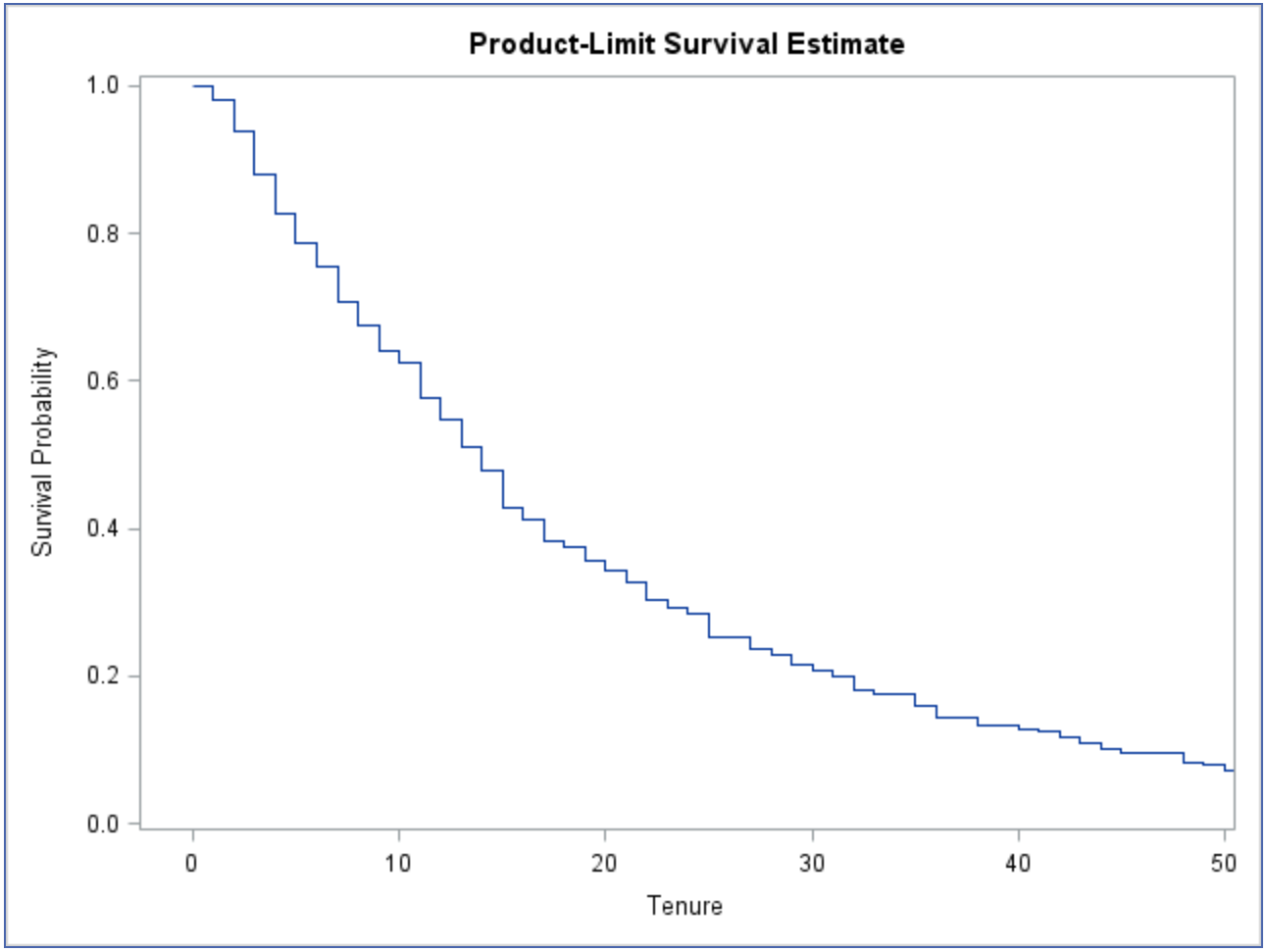
```
proc lifetest data=Survival.Loyalty;
     time Tenure;
run;

proc lifetest data=Survival.Loyalty maxtime=48;
     time Tenure;
run;
```

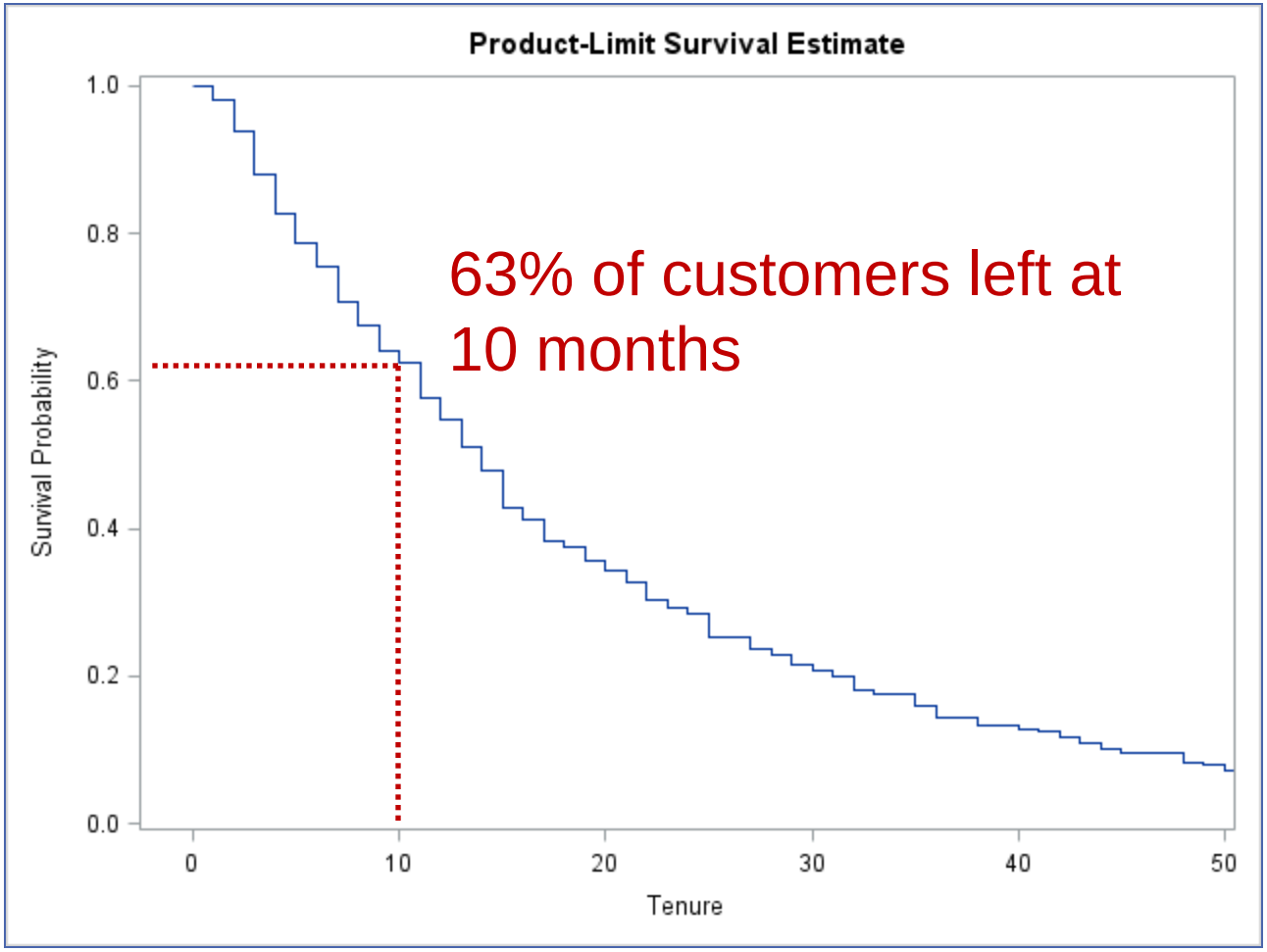
Survival Curves



Survival Curves



Survival Curves



Kaplan-Meier Method

- The Kaplan-Meier method existed long before Kaplan and Meier.
- Kaplan and Meier showed it was the maximum likelihood estimate for the nonparametric estimation of the survival curve.
- 3 situations for calculating survival curve:
  1. No censoring (very intuitive)
  2. Right Censored (somewhat intuitive)
  3. Complicated Censoring (not very intuitive)

Kaplan-Meier Method

- The Kaplan-Meier method existed long before Kaplan and Meier.
- Kaplan and Meier showed it was the maximum likelihood estimate for the nonparametric estimation of the survival curve.
- 3 situations for calculating survival curve:
  1. No censoring (very intuitive)

$$\hat{S}(t) = \frac{n_t}{N}$$

  2. Right Censored (somewhat intuitive)
  3. Complicated Censoring (not very intuitive)

Kaplan-Meier Method

- The Kaplan-Meier method existed long before Kaplan and Meier.
- Kaplan and Meier showed it was the maximum likelihood estimate for the nonparametric estimation of the survival curve.
- 3 situations for calculating survival curve:
  1. No censoring (very intuitive)

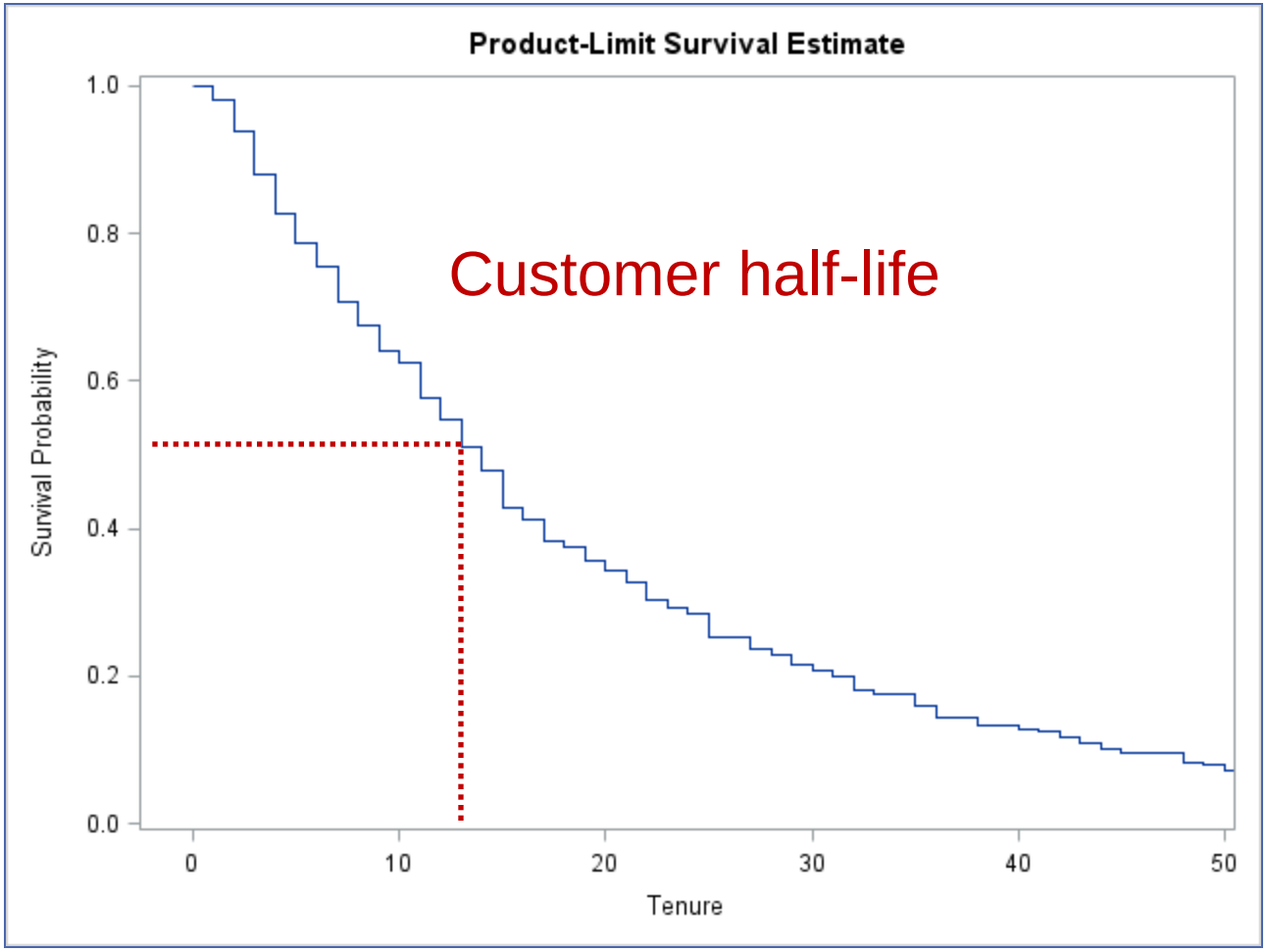
$$\hat{S}(t) = \frac{n_t}{N}$$

n_t → Individuals left at time t

N → Total individuals

  2. Right Censored (somewhat intuitive)
  3. Complicated Censoring (not very intuitive)

Median Customer Lifetime



Stratified Analysis

- Sometimes we want to compare different groups of observations with survival curves.
- What if we had different segments of our customers?
  - How do these segments react differently when it comes to customer retention (survival)?
- Different curves result in different models.
- SAS provides 3 tests that each have the same null hypothesis – all survival curves are **equal**.
  1. Log-rank test (developed by Mantel-Haenszel)
  2. Wilcoxon test
  3. Likelihood Ratio test (exponential distribution!)

Comparing Survival Curves

- Log-Rank test:

$$\text{LogRank} = \frac{1}{\hat{\sigma}^2}\left\{\sum_{j=1}^{r}\left(d_{1,j} - e_{1,j}\right)\right\}^2$$

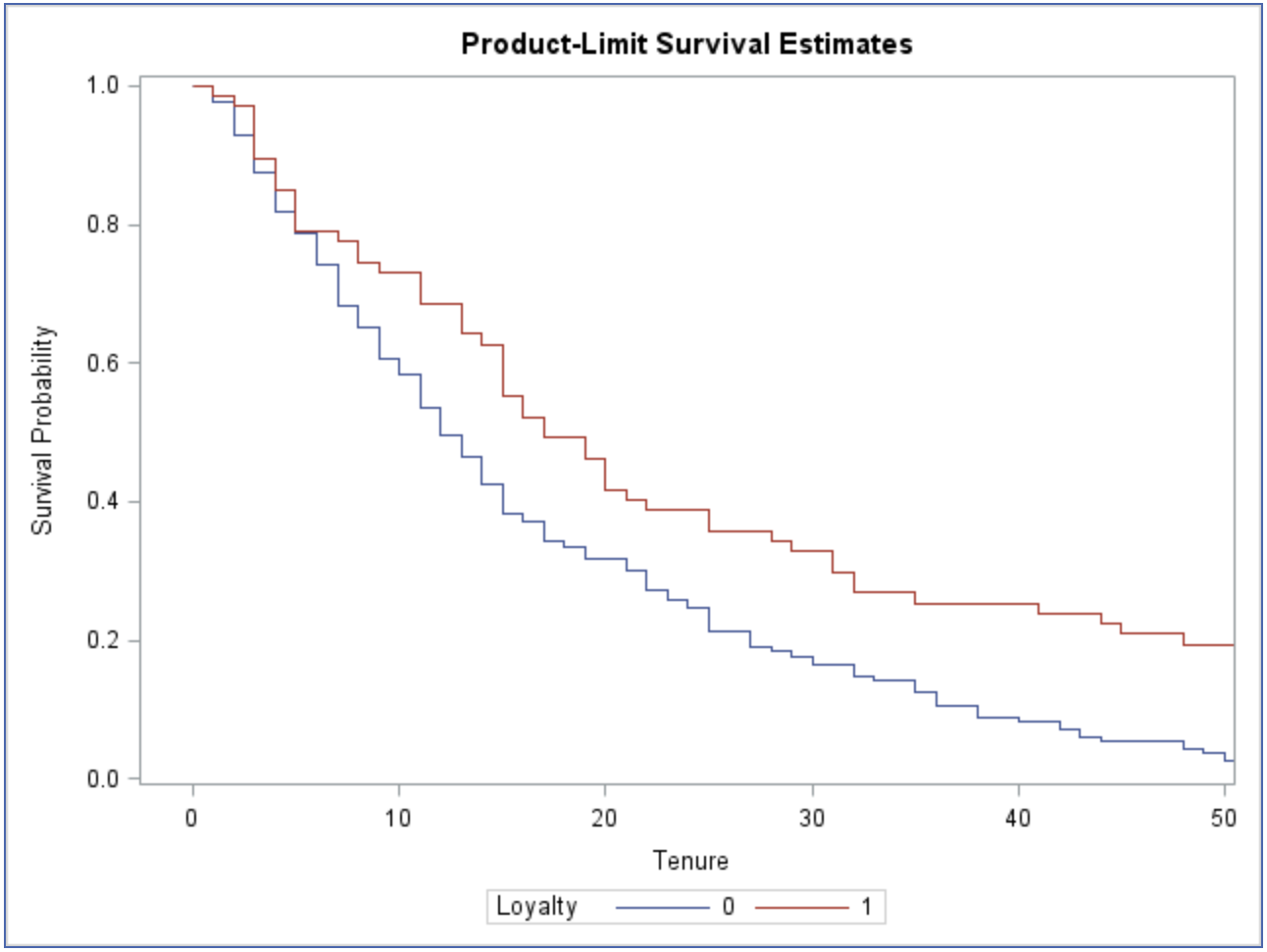
- Wilcoxon test (less sensitive to differences between groups later on in curve):

$$\text{Wilcoxon} = \frac{1}{\hat{\sigma}^2}\left\{\sum_{j=1}^{r}\left(d_{1,j} - e_{1,j}\right)n_j\right\}^2$$

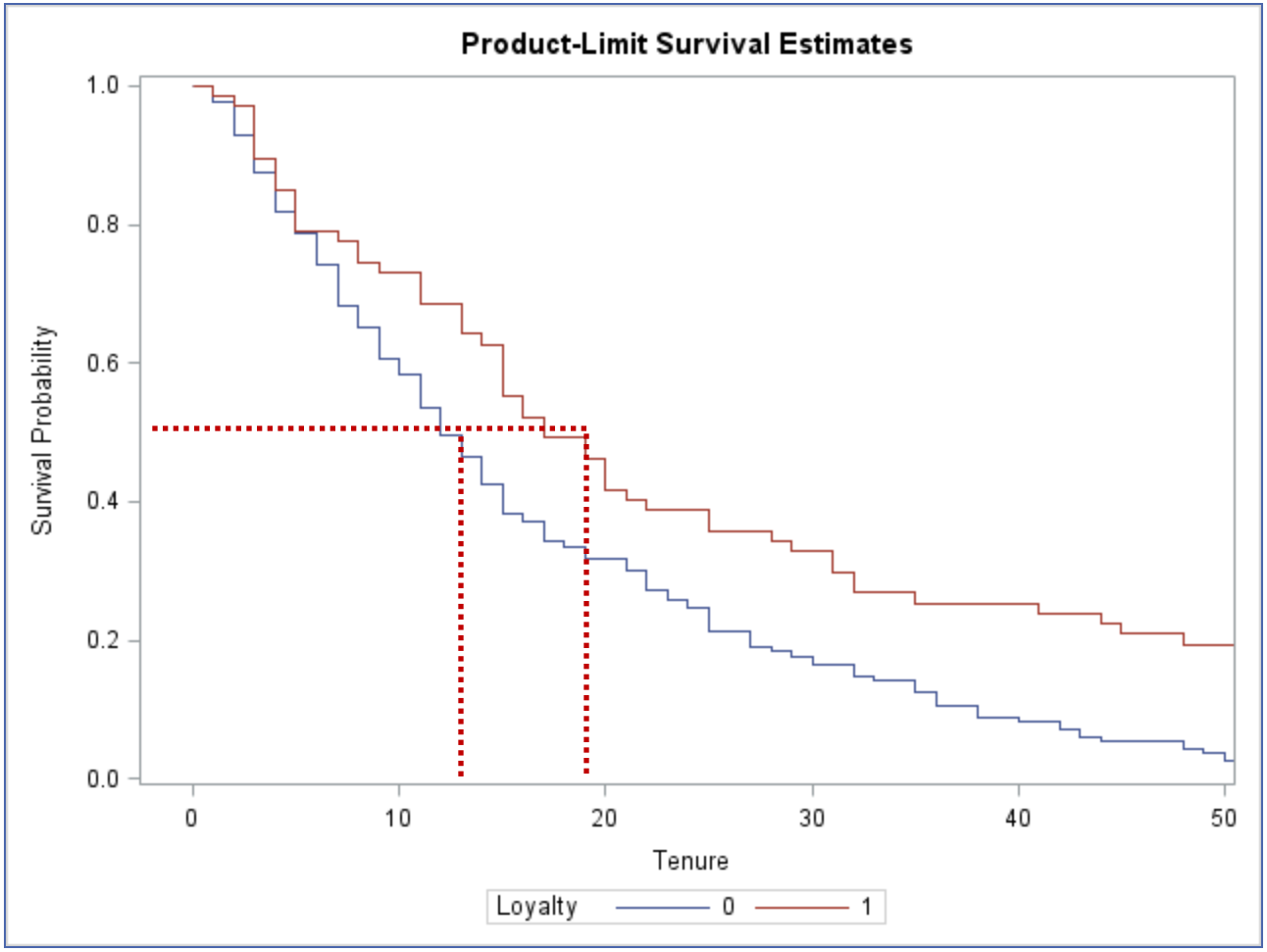
Stratified Analysis

```
proc lifetest data=Survival.Loyalty;
     time Tenure;
     strata Loyalty;
run;
```

Stratified Analysis



Stratified Analysis



LIFE-TABLE METHOD

Life-Table (Actuarial) Method

- Life-table method groups event times into intervals.
- Handles large data sets better than Kaplan-Meier.
- Life-table method allows for estimation of the *hazard function* in SAS.
- Calculation is based on conditional probabilities.

Stratified Analysis

```
proc lifetest data=Survival.Loyalty method=life;
     time Tenure;
     strata Loyalty;
run;
```

Prison Recidivism Data Set

- Maryland inmates released from prison.
- Week of next arrest for 432 released inmates (if under 52 weeks).
- Other recorded variables:
  - Financial Aid
  - Age at time of release
  - Race
  - Work experience
  - Marriage status
  - Parole
  - Number of prior convictions