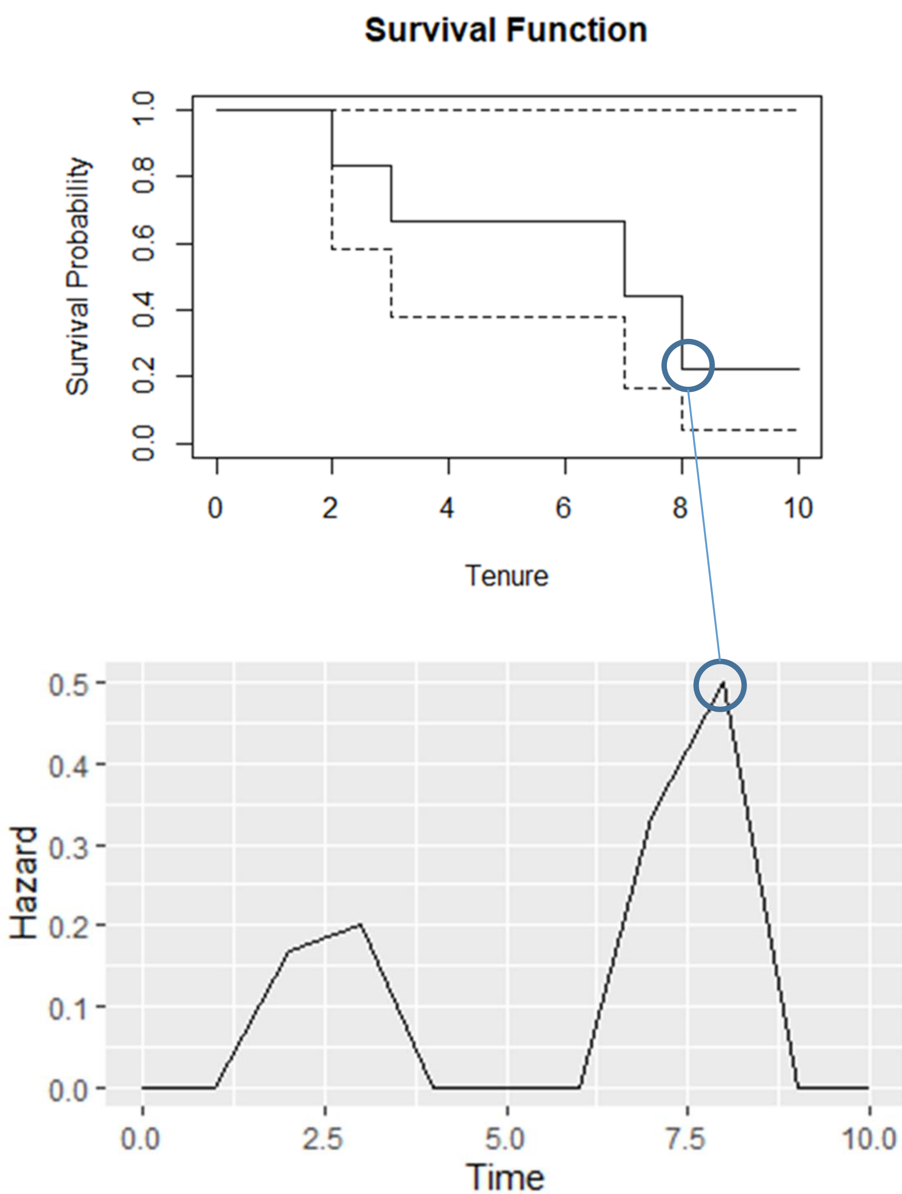
COX REGRESSION MODEL

PROPORTIONAL HAZARDS


Survival Function
Survival Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Tenure
Hazard
0.0
0.1
0.2
0.3
0.4
0.5
0.0
2.5
5.0
7.5
10.0
Time

AFT model:

- Recall the AFT model (model the time til event occurs…can get survival curves for each individual!):

$$T_i = e^{\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

- Then took the log to actually fit the model:

$$\log T_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

Proportional Hazards Model

- Alternative to modeling failure time is to model hazards.
- Hazard function is:

$$h(t) = e^{\beta_0+\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}} = e^{\beta_0} e^{\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}}$$

$$h(t) = h_0(t) e^{\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}}$$

- **Proportional hazard (Cox Regression) model:** model the log of the hazard directly:

$$\log h(t) = \log h_0(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

- Predictions shift the hazard rather than directly shifting the failure time like in the AFT model.

Proportional Hazards Model

- Alternative to modeling failure time is to model hazards.
- Hazard function is:

$$h(t) = h_0(t) e^{\beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

Baseline hazard function

- **Proportional hazard model:** model the log of the hazard directly:

$$\log h(t) = \log h_0(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

- Predictions shift the hazard rather than directly shifting the failure time like in the AFT model.

Proportional Hazards Model

- Alternative to modeling failure time is to model hazards.
- Hazard function is:

$$h(t) = h_0(t)e^{\beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

Predictors influencing hazard

- **Proportional hazard model:** model the log of the hazard directly:

$$\log h(t) = \log h_0(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

- Predictions shift the hazard rather than directly shifting the failure time like in the AFT model.

Proportional Hazards Model

- Why is the proportional hazard model so popular?
- Look at two different individuals x_i and x_j and their respective hazards:

$$h_i(t) = h_0(t)e^{\beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

$$h_j(t) = h_0(t)e^{\beta_1 x_{j,1} + \cdots + \beta_k x_{j,k}}$$

Proportional Hazards Model

- Why is the proportional hazard model so popular?
- Look at two different individuals x_i and x_j and their respective hazards:

$$h_i(t) = h_0(t)e^{\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}}$$

$$h_j(t) = h_0(t)e^{\beta_1 x_{j,1}+\cdots+\beta_k x_{j,k}}$$

$$\frac{h_i(t)}{h_j(t)} = e^{\beta_1(x_{i,1}-x_{j,1})+\cdots+\beta_k(x_{i,k}-x_{j,k})}$$

Proportional Hazards Model

- Why is the proportional hazard model so popular?
- Look at two different individuals x_i and x_j and their respective hazards:

$$h_i(t) = h_0(t)e^{\beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

$$h_j(t) = h_0(t)e^{\beta_1 x_{j,1} + \cdots + \beta_k x_{j,k}}$$

- **Hazard ratio** between the two:

$$\frac{h_i(t)}{h_j(t)} = e^{\beta_1 (x_{i,1} - x_{j,1}) + \cdots + \beta_k (x_{i,k} - x_{j,k})}$$

Proportional Hazards Model

- Why is the proportional hazard model so popular?
- Look at two different individuals x_i and x_j and their respective hazards:

$$h_i(t) = h_0(t)e^{\beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

$$h_j(t) = h_0(t)e^{\beta_1 x_{j,1} + \cdots + \beta_k x_{j,k}}$$

- **Hazard ratio** between the two:

$$\frac{h_i(t)}{h_j(t)} = \boxed{e^{\beta_1 (x_{i,1} - x_{j,1}) + \cdots + \beta_k (x_{i,k} - x_{j,k})}}$$

No longer depends on time!
Constant **proportion** on **hazards**.

Accelerated Failure Time Model

Initial

- Investigate Survival Curves
- Investigate Hazard Function

Distributions

- Find “best” distribution
  - Graphical methods
  - Statistical Tests

Model Building

- Select significant variables
- Finalize model

Accelerated Failure Time Model

Initial

- Investigate Survival Curves
- Investigate Hazard Function

Distributions

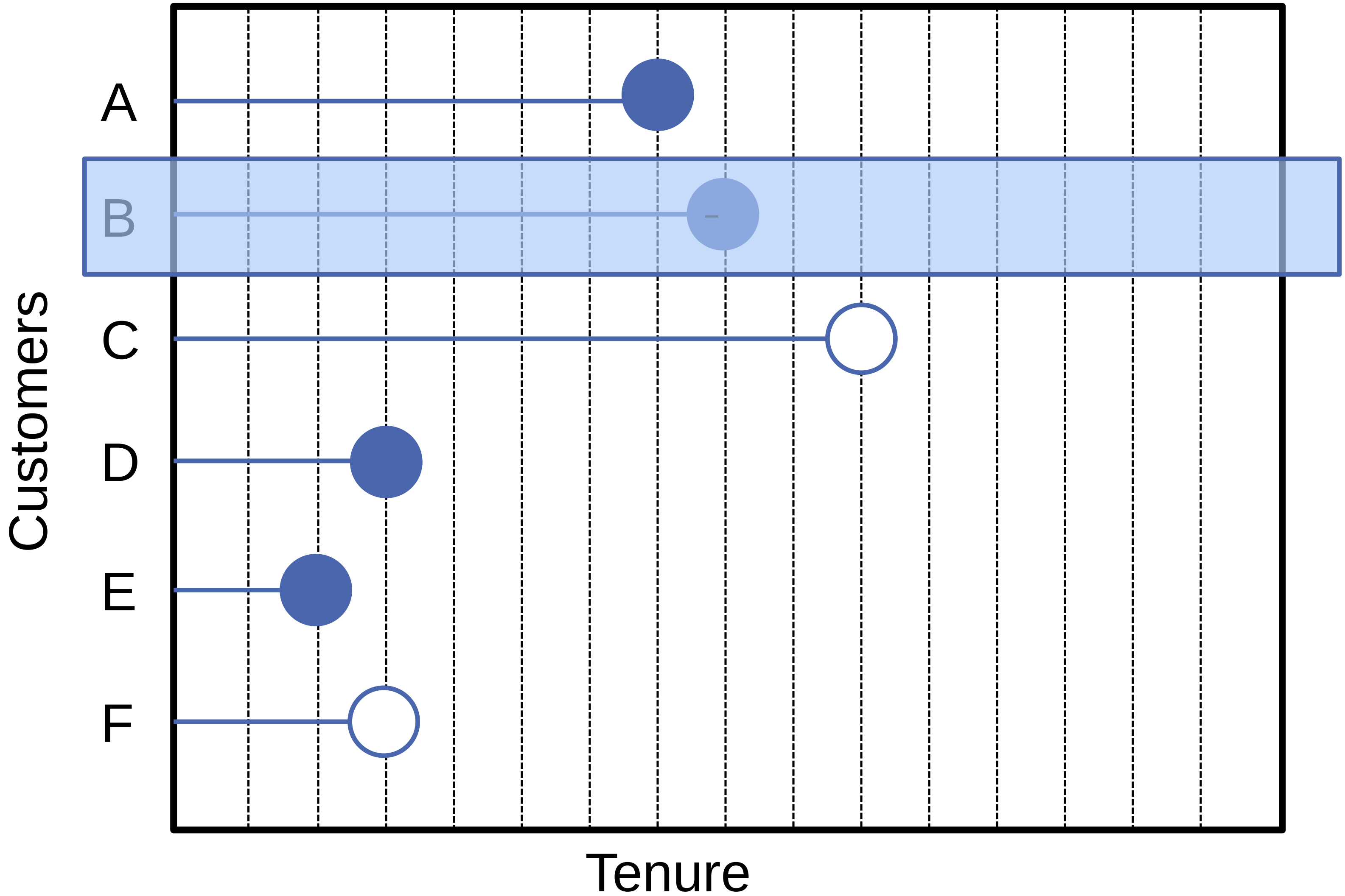
- Find “best” distribution
  - Graphical methods
  - Statistical Tests

Model Building

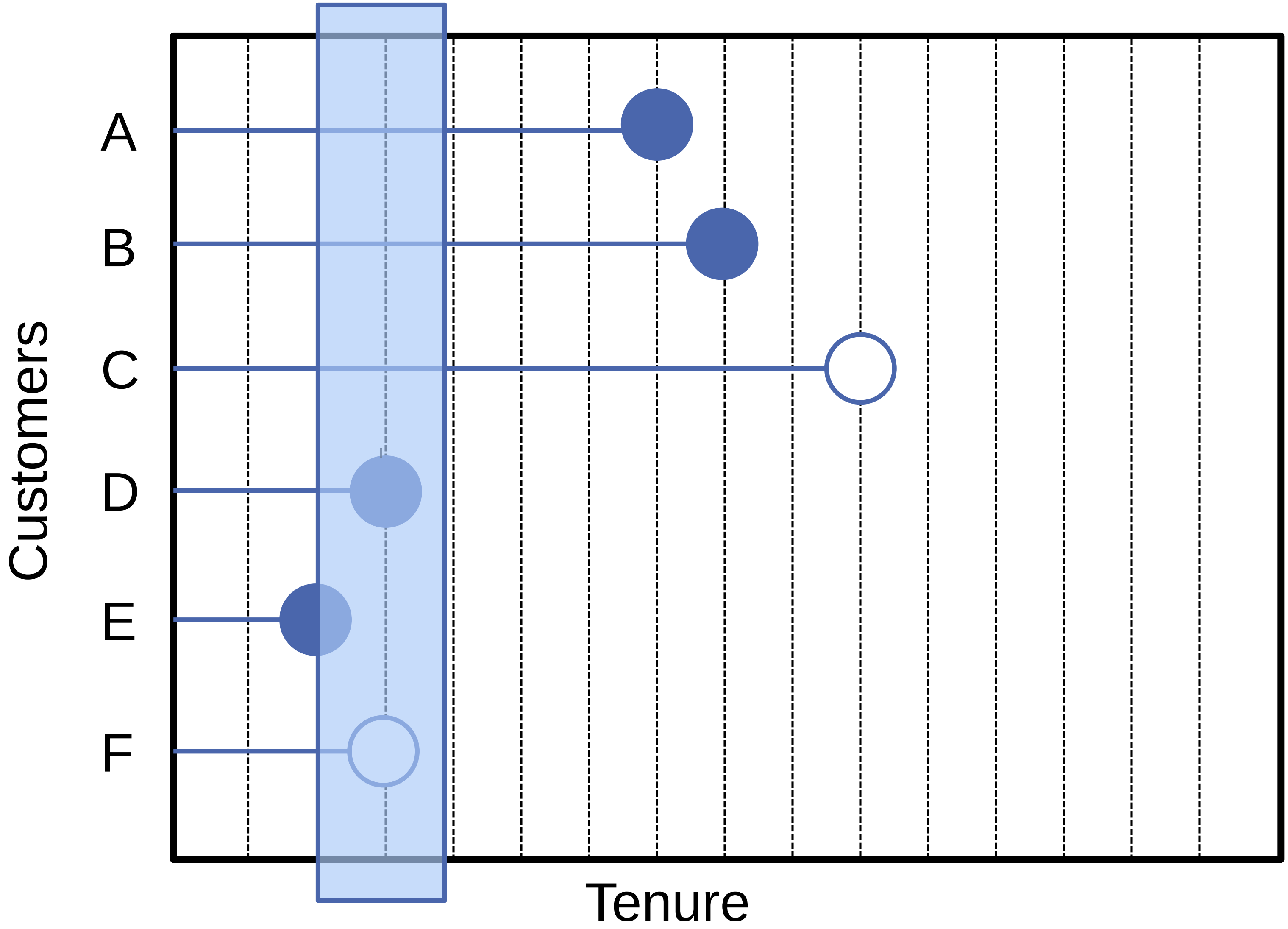
- Select significant variables
- Finalize model

Proportional Hazards changes this!!!

Accelerated Failure Time Model



Proportional Hazards Model

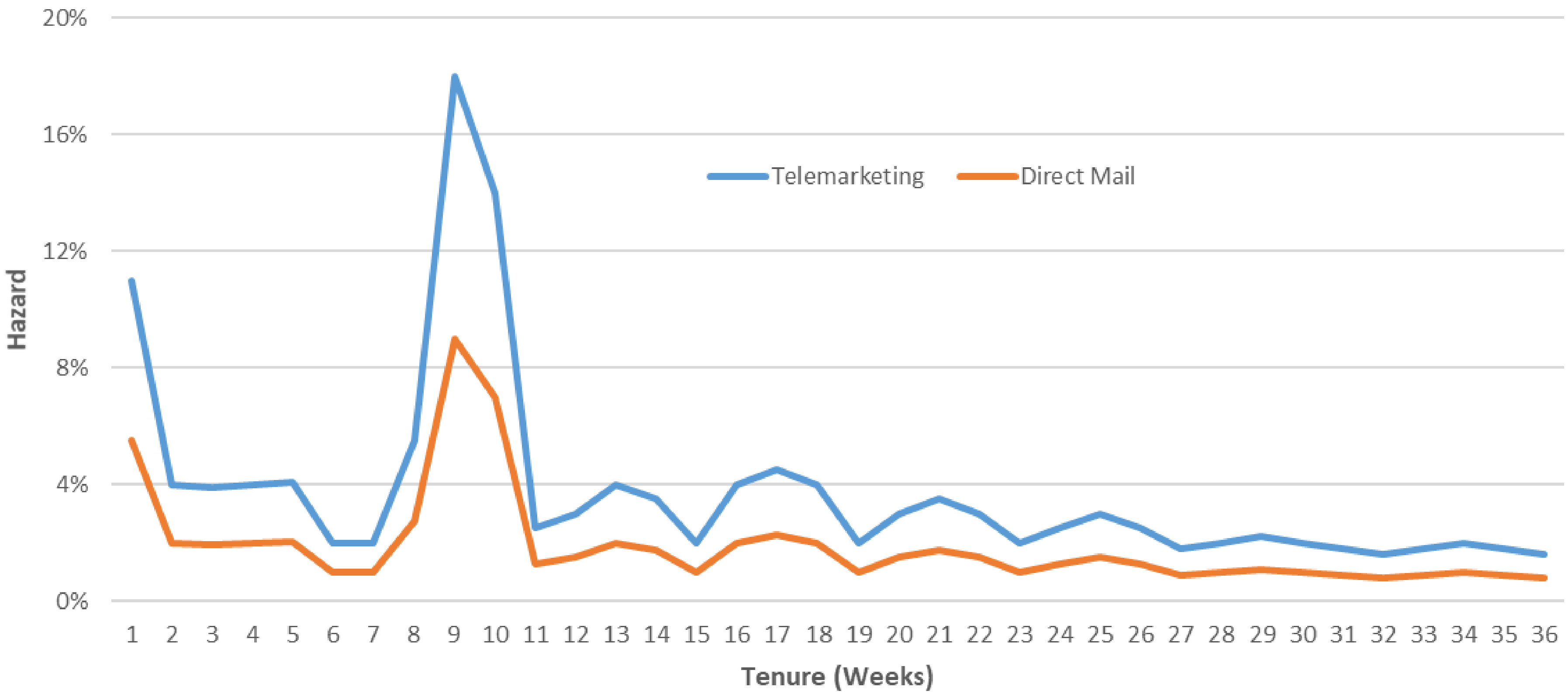


PH Model – Example

- “On average, a customer who signed up via direct mail stays twice as long compared to a customer who signed up via telemarketing.”

- Results do not say how long someone will last, only relative length of tenure between two groups.
- Assume that factors measured at an initial time point have a uniform proportional effect on hazards between individuals (or groups).

PH Model – Example



AFT vs. PH Models

- **AFT Model:** Predictors have a multiplicative effect on failure time:

$$T_i = e^{\beta_0+\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}} = e^{\beta_0} e^{\beta_1 x_{i,1}+\cdots\beta_k x_{i,k}}$$

$$T_i = T_0 \boxed{e^{\beta_1 x_{i,1}+\cdots\beta_k x_{i,k}}}$$

- **PH Model:** Predictors have a multiplicative effect on hazard:

$$h(t) = h_0(t) \boxed{e^{\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}}}$$

Weibull Distribution!

- Weibull (and Exponential) model is a rare case where there is a relationship between the two models:

$$T_i = e^{\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}}$$

$$\tilde{\beta}_j = \frac{-\beta_j}{\sigma}$$

$$h(t) = h_0(t) e^{\tilde{\beta}_1 x_{i,1} + \cdots + \tilde{\beta}_k x_{i,k}}$$

Proportional Hazards Model – R

```r
recid.ph <- coxph(Surv(week, arrest) ~ fin + age  +
                  wexp + mar +paro + prio, data = recid)

summary(recid.ph)
```

Proportional Hazards Model – R

```
 n= 432, number of events= 114

         coef exp(coef) se(coef)      z Pr(>|z|)
fin  -0.36554   0.69382  0.19090 -1.915  0.05552 .
age  -0.05633   0.94523  0.02189 -2.573  0.01007 *
wexp -0.15699   0.85471  0.21208 -0.740  0.45916
mar  -0.47130   0.62419  0.38027 -1.239  0.21520
paro -0.07792   0.92504  0.19530 -0.399  0.68991
prio  0.08966   1.09380  0.02871  3.123  0.00179 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Proportional Hazards Model – R

```
     exp(coef) exp(-coef) lower .95 upper .95
fin     0.6938     1.4413    0.4773    1.0087
age     0.9452     1.0579    0.9055    0.9867
wexp    0.8547     1.1700    0.5640    1.2952
mar     0.6242     1.6021    0.2962    1.3152
paro    0.9250     1.0810    0.6308    1.3564
prio    1.0938     0.9142    1.0340    1.1571

Concordance= 0.639  (se = 0.027 )
Likelihood ratio test= 32.14  on 6 df,   p=2e-05
Wald test            = 30.79  on 6 df,   p=3e-05
```

Hazard Ratio

- If a parameter estimate is positive, increases in that variable increase the expected hazard.
  - **Increase** the rate/risk of failure
- If a parameter estimate is negative, increases in that variable decrease expected hazard.
  - **Decrease** in the rate/risk of failure
- $100 \times \left(e^{\beta} - 1\right)$ is the % increase in the expected hazard for each one-unit increase in the variable.
- e^{β} is the hazard ratio – the ratio of the hazards for each one-unit increase in the variable.

Recidivism Parameter Interpretation

Variable	β Estimate	$100(e^{\beta} - 1)$
Financial Aid	-0.347	-29.3%
Age at Release	-0.067	-6.5%
Prior Convictions	0.097	10.2%

These parameter estimates are from the model with only Financial Aid, Age at Release and Prior Convictions

Recidivism Parameter Interpretation

Variable	β Estimate	$100\left(e^{\beta}-1\right)$
Financial Aid	-0.347	-29.3%
Age at Release	-0.067	-6.5%
Prior Convictions	0.097	10.2%

For those who received financial aid, the rate of recidivism decreased by 29.3% compared to those who did not receive financial aid, holding all other variables constant.

Recidivism Parameter Interpretation

Variable	β Estimate	$100(e^{\beta} - 1)$
Financial Aid	-0.347	-29.3%
Age at Release	-0.067	-6.5%
Prior Convictions	0.097	10.2%

For every year older at the time of release, the rate of recidivism decreases by 6.5%, holding all other variables constant.

Recidivism Parameter Interpretation

Variable	β Estimate	$100\left(e^{\beta}-1\right)$
Financial Aid	-0.347	-29.3%
Age at Release	-0.067	-6.5%
Prior Convictions	0.097	10.2%

For every increase in prior convictions, the rate of recidivism increases by 10.2%, holding all other variables constant.

ESTIMATION

Semiparametric Models

- In AFT and PH models, estimation depends on some distributional assumption around either the failure time or the baseline hazard.
- However, in PH models, Cox noticed that the likelihood can be split into two pieces:
  - 1st piece: depends on $h_0(t)$ and the parameters
    - Treat as non-parametric (no assumptions about form or distribution)
  - 2nd piece: **only** depends on the parameters
    - Treat as parametric (know the form)
- This is why it is called a **semiparametric** model.

Cox Regression Model

- Using the semiparametric model approach, we can basically ignore ever estimating anything about the baseline hazard $h_0(t)$ – the **Cox regression model**.

- Basically, Cox disregarded the first piece of the likelihood and maximized the second piece – still a PH model.

Partial Likelihood Estimation

- This is the more important piece of the work done by Sir David Cox in his original article.
- Estimates are obtained by maximizing the **partial likelihood** – only one piece that depends on the predictors, not the entire thing.
  - All we care about is ratios between hazards.

OPTIONAL: Too Much Info on PMLE

- Since estimation for Cox regression models hazards (at each time point), if more than one event occurs at a given time point, there is a tie.
- Common methods to construct an appropriate partial likelihood for breaking ties: Efron (R default), Breslow (SAS default), exact
- Safe to go with Efron because it does better for higher numbers of ties.

Partial Likelihood Downfalls

- Some information about the parameters is lost due to the partial likelihood estimation – inefficient estimates.
- Inefficiency is rather small.
- Estimates still have some desired properties:
  - Unbiased
  - Estimates can be tested in the same way as before.

Comparative Risks

- Cox regression essentially is estimating a subject's **relative** likelihood of failure at a specific time compared to everyone else in the risk set at that time.
  - Normal people words example: Conditional on a failure happening at time t, how likely was it to happen to subject *i* out of everyone remaining at that time?

- Any estimation/inference (coefficients, hazard ratios, etc.) is still valid, but contrary to the AFT, Cox regression model **DO NOT** make any absolute predictions of time or risk.

Assumptions

- Wait…!?!?!?! I thought you said there were no distributional assumptions!
- Still other assumptions we need to check:
  - Linearity
  - Proportional hazards (no interactions with time)

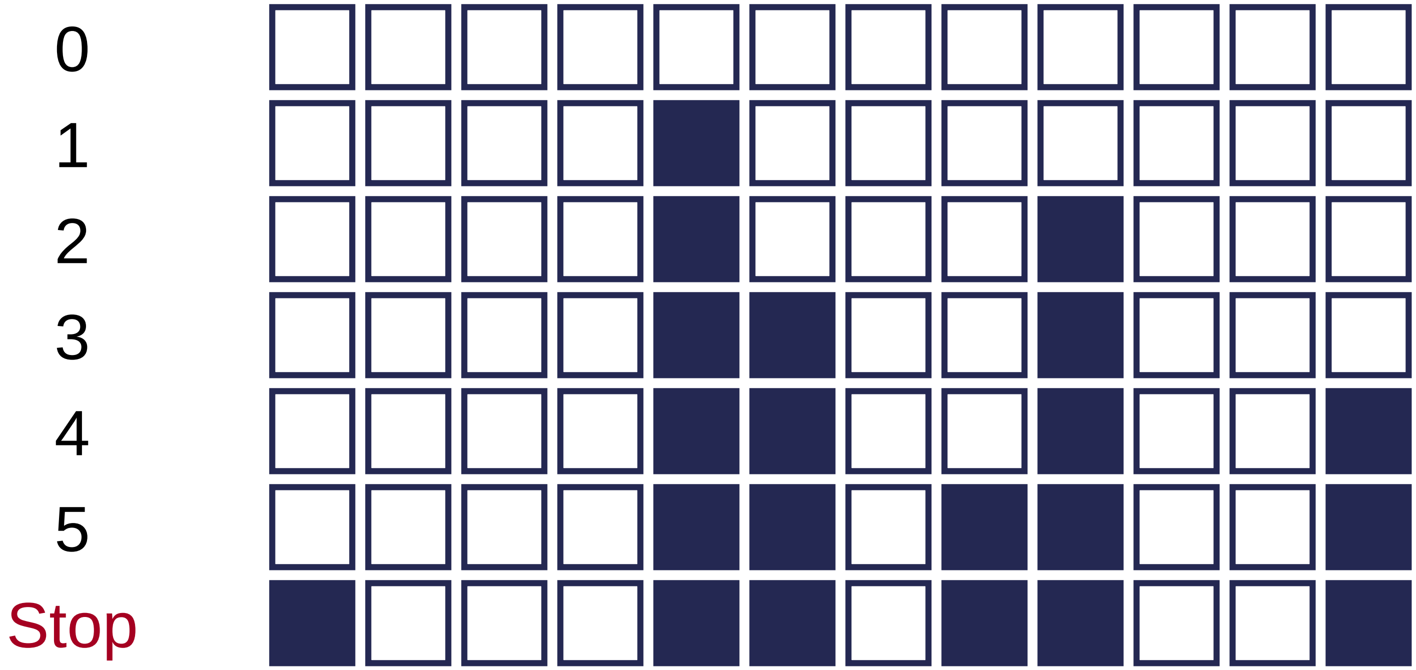
- Will deal with these later…

AUTOMATIC SELECTION TECHNIQUES

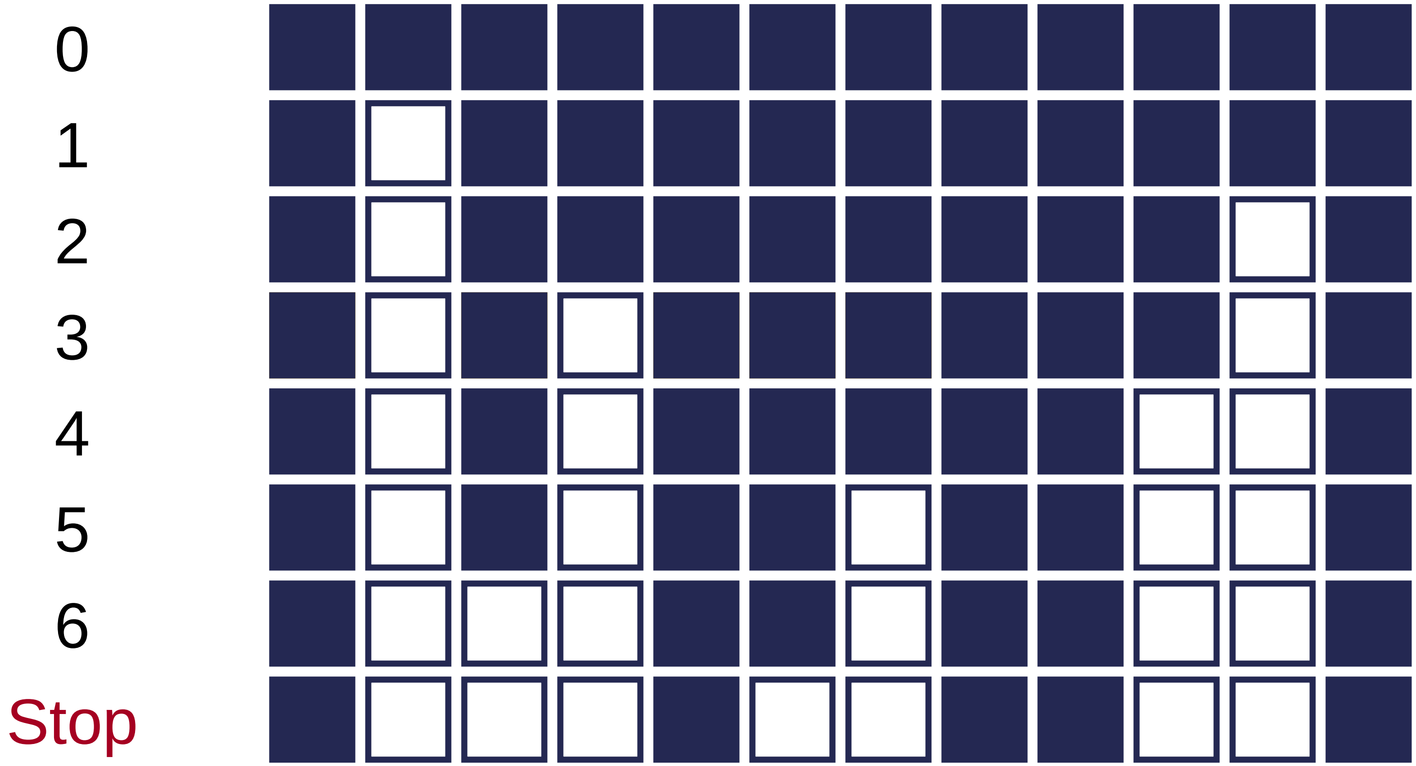
Automatic Selection Techniques

- One of the benefits of PROC PHREG is the automatic selection techniques that it employs.
- Has similar selection techniques as PROC LOGISTIC:
  - Forward
  - Backward
  - Stepwise

Forward Selection

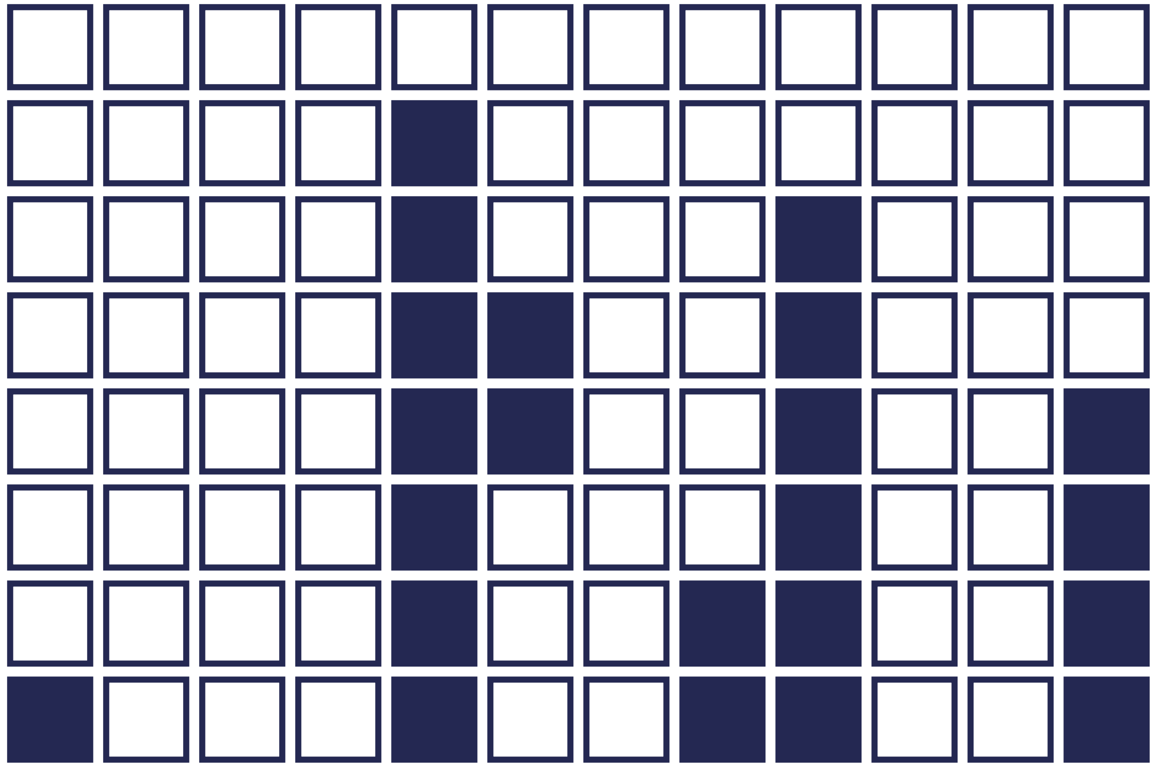


Backward Elimination



Stepwise Selection

0
1
2
3
4
5
6
Stop



Automatic Selection Techniques – R

```
step.model <- step(empty.model,
                   scope = list(lower=formula(empty.model),
                                upper=formula(full.model)),
                   direction = "both")
summary(step.model)
```

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
age	-0.06042	0.94137	0.02085	-2.897	0.00376	**
prio	0.09751	1.10243	0.02722	3.583	0.00034	***
fin	-0.36020	0.69753	0.19049	-1.891	0.05864	.
mar	-0.53312	0.58677	0.37276	-1.430	0.15266	

PREDICTIONS

Estimating Survival Curves

- Once we've obtained parameter estimates from the partial likelihood, we can plug it into the "full likelihood" and nonparametrically estimate the remaining piece.
  - Think combining partial MLE and Kaplan-Meier…

- Now we can estimate survival curves for predefined predictor values (combinations of the x's).

Estimated Survival Curves – R

```
newdata <- data.frame(fin = c(1, 0), age = 30,
                      wexp = c(1, 0), mar = 0, paro = 0,
                      prio = c(0, 4))

ggsurvplot(survfit(recid.ph, newdata), data = newdata,
           break.y.by = 0.1, palette = c("purple", "black"),
           ylab = "Survival Probability", xlab = "week",
           legend.labs = c("1", "2"), legend.title = "subject")
```

Estimated Survival Curves – R