

EAS509 Homework 6 (50 points).

Submit your answers as a single pdf attach all R code. Failure to do so will result in grade reduction.

Question 1 (25 points)

For each question, state whether or not the censoring mechanism is independent. Justify your answer with a short statement. (5 points for each)

a) In a study of disease relapse, due to a careless research scientist, all patients whose phone numbers begin with the number “2” are lost to follow up.

There are 2 types of scenarios we can think of:

- 1) Censoring mechanism is independent if we assume that phone number digits have no relationship with diseases.
- 2) Censoring mechanism is dependent because depending on phone number we can get area code or geographical location where some diseases will occur in specific areas which in return those areas can be also identified by phone codess.

b) In a study of longevity, a formatting error causes all patient ages that exceed 99 years to be lost (i.e. we know that those patients are more than 99 years old, but we do not know their exact ages).

Censoring mechanism is independent as we will not considr anybody at all because If We remove all people who lived more than 99 and died and people who survived for short time after 99 there wont be considered or left to study

c) Hospital A conducts a study of longevity. However, very sick patients tend to be transferred to Hospital B, and are lost to follow up.

Censoring mechanism is very much dependent here, as there are 2 types of hospitals for types of diseases. Hospital B will get more sick patients with dangeours dieases and also as it has facilities more man power to treat such diseases when compared to hospital A. Also, when we consider the morality rate Hospital B will have more because of treating such deadly diseases.

d) In a study of unemployment duration, the people who find work earlier are less motivated to stay in touch with study investigators, and therefore are more likely to be lost to follow up.

Censoring mechanism is dependent here, as people will ofcourse dropout if they land a job much earlier before the investigation is finished.

e) In a study of pregnancy duration, women who deliver their babies pre-term are more likely to do so away from their usual hospital, and thus are more likely to be censored, relative to women who deliver full-term babies.

Censoring mechanism is dependent here, as women willn't here visit the hospital as babies are born earlier than regular estimated time..

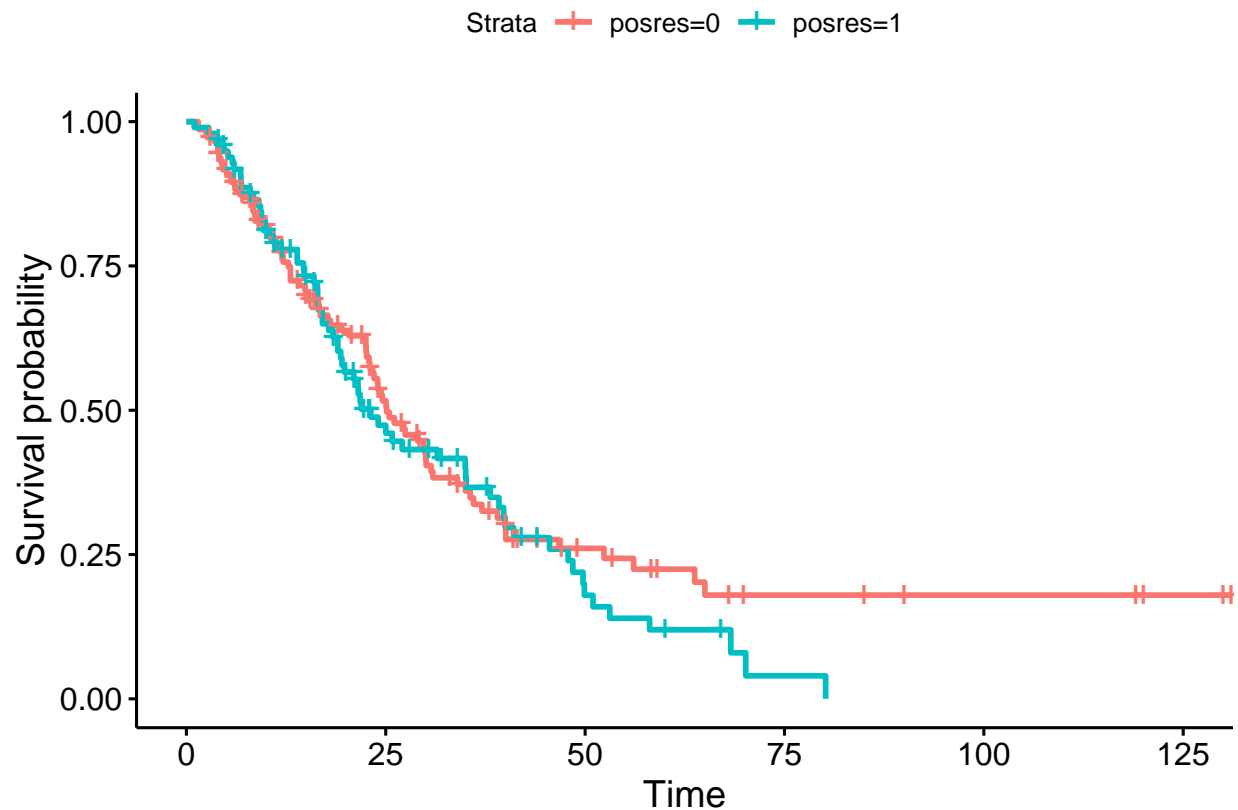
Question 2 (25 points)

A data set from "DATA.csv" represents publication times for 244 clinical trials funded by the National Heart, Lung, and Blood Institute. Using Log-Rank Test in R, estimate if the Kaplan-Meier Survival Curves from two subpopulations stratified by "posres" variable are significantly different.

```
clinic_data <- read.csv("DATA.csv")
head(clinic_data)
```

##	posres	multi	clinend	mech	sampsize	budget	impact	time	status
## 1	0	0	1	R01	39876	8.016941	44.016	11.203285	1
## 2	0	0	1	R01	39876	8.016941	23.494	15.178645	1
## 3	0	0	1	R01	8171	7.612606	8.391	24.410678	1
## 4	0	0	1	Contract	24335	11.771928	15.402	2.595483	1
## 5	0	0	1	Contract	33357	76.517537	16.783	8.607803	1
## 6	0	0	1	Contract	10355	9.809938	16.783	8.607803	1

```
sm <- survfit(Surv(time, status) ~ posres, data = clinic_data)
ggsurvplot(sm, data=clinic_data)
```



```
lrt <- survdiff(Surv(time, status) ~ posres, data = clinic_data)
lrt
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ posres, data = clinic_data)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## posres=0 146      87     92.6    0.341    0.844
## posres=1  98      69     63.4    0.498    0.844
##
## Chisq= 0.8  on 1 degrees of freedom, p= 0.4
```

We can observe that output of P-value is $0.4 > 0.05$ (default value). With this we don't have enough evidence to reject the null hypothesis. So, I can say that NO Significant Difference is observed between Kaplan-Meier Survival Curves from two subpopulations stratified by "posres" variable .