PHC 6000: Homework 1

Joe Brew

January 20, 2014

UF-ID: 0402-8902

Contents

Problem 1: Incidence versus Prevalence

A

Point prevalence.

B

Since we dont have exact information on each person **incidence rate** best describes this fraction. Incidence rate addresses the fact that we don't have specific information on all population members and the fact that there is a time element (cases per person-time units).

C

Assuming that the "numer of cervical cancer cases in Florida in 2009" refers both to those which were prevalent at the beginning of the year as well as incident 2009 cases, then the correct answer is **period prevalence**.

Problem 2: Survival analysis assumptions

Differences

Two differences between the actuarial / life table approach and the Kaplan-Meier method are:

1. The actuarial / life table method requires a priori knowledge of time intervals to be used (the Kaplan-Meier method creates intervals as a function of actual event/censorship times).
2. In the actuarial / life table method (unlike Kaplan-Meier), it is not necessary to know the exact time of event or censorship.

Shared assumptions

Both the actuarial / life table method as well as the Kaplan-Meier method assume that :

1. There are no secular trends (ie, big changes from outside influences).
2. The censored population are no different (in terms of probability of having event) from the non-censored population (ie, independence of event and censoring.)

Problem 3: Cumulative incidence

Table 1: Life Table

	interval	casesAtStart	deathsDuring	withdrawsDuring	casesAtRisk	condProbOfEvent	condProbOfSurv	cumProbOfSurv	cumProbOfEvent
1	0-5	0.00	2.00	1.00	10.00	0.20	0.80	0.80	0.20
2	6-10	3.00	2.00	1.00	7.00	0.29	0.71	0.57	0.43
3	11-15	6.00	3.00	0.00	4.00	0.75	0.25	0.18	0.82
4	16-20	9.00	1.00	0.00	1.00	1.00	0.00	0.00	1.00

[1] ''

1

The cumulative probability of death over the 20 months period using the life table interval approach is 1.

2

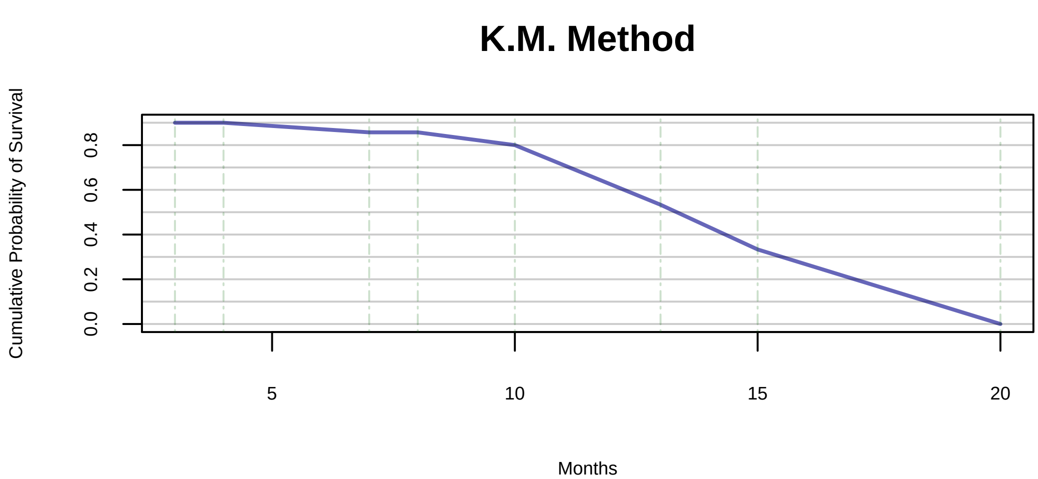
The cumulative probability of survival over the 20 months period using the life table interval approach is 0.

3

The Kaplan-Meier approach, unlike the Life Table approach, relies on setting up time intervals as a function of the actual time of event / withdrawal. See below table:

Table 2: Kaplan-Meier Table

	interval	casesAtStart	deathsDuring	withdrawsDuring	casesAtRisk	condProbOfEvent	condProbOfSurv	cumProbOfSurv	cumProbOfEvent
1	3.00	0.00	1.00	0.00	10.00	0.10	0.90	0.90	0.10
2	4.00	1.00	0.00	1.00	9.00	0.00	1.00	0.90	0.10
3	7.00	3.00	1.00	0.00	7.00	0.14	0.86	0.86	0.14
4	8.00	4.00	0.00	1.00	6.00	0.00	1.00	0.86	0.14
5	10.00	5.00	1.00	0.00	5.00	0.20	0.80	0.80	0.20
6	13.00	7.00	1.00	0.00	3.00	0.33	0.67	0.53	0.47
7	15.00	8.00	1.00	0.00	2.00	0.50	0.50	0.33	0.67
8	20.00	9.00	1.00	0.00	1.00	1.00	0.00	0.00	1.00



[1] ''

Problem 4: Incidence density - individual data

Total number of person-years

The total number of person-years contributed is equal to the sum of each participants' time before censoring or onset of disease.

```
> answer <- 3 + 5 + 2.4 + 3.6 + 4 + 1.9 + 4.1 + 2 + 3 + 5
```

Answer: **34 person-years**

5-year incidence density

The incidence density is a fraction which contains the number of incident cases in (numerator) over the sum of the person-time of all individuals at risk (denominator).

```
> personNum <- 1:10
> status <- c("d", "c", "c", "d", "c", "c", "c", "d", "c", "c")
> interval <- c(3,5,2.4,3.6,4,1.9,4.1,2,3,5)
> prob4 <- as.data.frame(cbind(personNum, status, interval))
> answer <- nrow(prob4[which(prob4$status=="d"),]) / sum(interval)
```

In this case it is equal to **0.088** new cases per person-year (work shown above).

Problem 5: Incidence rate - group data

Using the midpoint estimation method to calculate the incidence rate of X assumes a normal distribution of removal from study population time (either by censoring or event). This assumption allows for us to simply take the (initial population plus the end population) divided by two. Once the midpoint population has been estimated, the incidence rate is simply the number of incident cases over the midpoint population (for a given set of time - in this case, two years):

```
> initialPop <- 300
> newCases <- 5
> censored <- 35
> endPop <- initialPop - newCases - censored
> midPointPop <- (initialPop + endPop) / 2
> incidenceRate <- newCases / midPointPop
```

0.018 incident cases for every 2 person-years (equivalent to 0.009 per person-year).

Details

Problem 3 (work shown)

```
> library(xtable)
> personNum <- as.numeric(c(1:10))
> months <- as.numeric(as.character(c(10, 12, 4, 3, 7, 5, 8, 15, 13, 20)))
> status <- c("d", "d", "c", "d", "d", "d", "c", "d", "d", "d")
> prob3 <- as.data.frame(cbind(personNum, months, status),
+                        stringsAsFactors = FALSE)
> prob3$months <- as.numeric(as.character(prob3$months))
> prob3$personNum <- as.numeric(as.character(prob3$personNum))
> prob3$interval <- ifelse(prob3$months <= 5,
+                          "0-5",
+                          ifelse(prob3$months >=6 &
+                                   prob3$months <=10,
+                                 "6-10",
+                                 ifelse(prob3$months >=11 &
+                                          prob3$months <=15,
+                                        "11-15",
+                                        ifelse(prob3$months >=16 &
+                                                 prob3$months <=20,
+                                               "16-20",
+                                               NA))))
> interval <- c("0-5", "6-10", "11-15", "16-20")
> casesAtStart <- c(NA, NA, NA, NA)
> deathsDuring <- c(NA, NA, NA, NA)
> withdrawsDuring <- c(NA, NA, NA, NA)
> casesAtRisk <- c(NA, NA, NA, NA)
> condProbOfEvent <- c(NA, NA, NA, NA)
> condProbOfSurv <- c(NA, NA, NA, NA)
> cumProbOfSurv <- c(NA, NA, NA, NA)
> cumProbOfEvent <- c(NA, NA, NA, NA)
> lifeTable <- as.data.frame(cbind(interval, casesAtStart, deathsDuring,
+                                  withdrawsDuring, casesAtRisk,
+                                  condProbOfEvent, condProbOfSurv,
+                                  cumProbOfSurv, cumProbOfEvent))
> #casesAtStart
> lifeTable$casesAtStart <- as.numeric(lifeTable$casesAtStart)
> lifeTable$casesAtStart[which(lifeTable$interval== "0-5")] <- 0
> lifeTable$casesAtStart[which(lifeTable$interval== "6-10")] <-
+   nrow(prob3[which(prob3$months <= 5),])
> lifeTable$casesAtStart[which(lifeTable$interval== "11-15")] <-
+   nrow(prob3[which(prob3$months <= 10),])
> lifeTable$casesAtStart[which(lifeTable$interval== "16-20")] <-
+   nrow(prob3[which(prob3$months <= 15),])
> #deathsDuring
> lifeTable$deathsDuring <- as.numeric(lifeTable$deathsDuring)
> lifeTable$deathsDuring[which(lifeTable$interval == "0-5")] <-
+   nrow(prob3[which(prob3$status=="d"
+                  & prob3$months >=0
+                  & prob3$months <=5),])
> lifeTable$deathsDuring[which(lifeTable$interval == "6-10")] <-
+   nrow(prob3[which(prob3$status=="d"
+                  & prob3$months >=6
+                  & prob3$months <=10),])
> lifeTable$deathsDuring[which(lifeTable$interval == "11-15")] <-
+   nrow(prob3[which(prob3$status=="d"
+                  & prob3$months >=11
+                  & prob3$months <=15),])
> lifeTable$deathsDuring[which(lifeTable$interval == "16-20")] <-
+   nrow(prob3[which(prob3$status=="d"
+                  & prob3$months >=16
+                  & prob3$months <=20),])
> #withdrawsDuring
> lifeTable$withdrawsDuring <- as.numeric(lifeTable$withdrawsDuring)
> lifeTable$withdrawsDuring[which(lifeTable$interval == "0-5")] <-
+   nrow(prob3[which(prob3$status=="c"
+                  & prob3$months >=0
+                  & prob3$months <=5),])
> lifeTable$withdrawsDuring[which(lifeTable$interval == "6-10")] <-
+   nrow(prob3[which(prob3$status=="c"
+                  & prob3$months >=6
+                  & prob3$months <=10),])
> lifeTable$withdrawsDuring[which(lifeTable$interval == "11-15")] <-
+   nrow(prob3[which(prob3$status=="c"
+                  & prob3$months >=11
+                  & prob3$months <=15),])
> lifeTable$withdrawsDuring[which(lifeTable$interval == "16-20")] <-
+   nrow(prob3[which(prob3$status=="c"
+                  & prob3$months >=16
+                  & prob3$months <=20),])
> #casesAtRisk
> lifeTable$casesAtRisk <- as.numeric(lifeTable$casesAtRisk)
> lifeTable$casesAtRisk[which(lifeTable$interval== "0-5")] <- 10
> lifeTable$casesAtRisk[which(lifeTable$interval== "6-10")] <-
+   nrow(prob3) - nrow(prob3[which(prob3$months <= 5),])
> lifeTable$casesAtRisk[which(lifeTable$interval== "11-15")] <-
+   nrow(prob3) - nrow(prob3[which(prob3$months <= 10),])
> lifeTable$casesAtRisk[which(lifeTable$interval== "16-20")] <-
+   nrow(prob3) - nrow(prob3[which(prob3$months <= 15),])
> #condProbOfEvent
> lifeTable$condProbOfEvent <- as.numeric(lifeTable$condProbOfEvent)
> lifeTable$condProbOfEvent <- lifeTable$deathsDuring / lifeTable$casesAtRisk
> #condProbOfSurv
> lifeTable$condProbOfSurv <- as.numeric(lifeTable$condProbOfSurv)
> lifeTable$condProbOfSurv <- 1- lifeTable$condProbOfEvent
```

```
> #cumProbOfSurv
> lifeTable$cumProbOfSurv <- as.numeric(lifeTable$cumProbOfSurv)
> lifeTable$cumProbOfSurv[1] <- lifeTable$condProbOfSurv[1]
> for (i in 2:4){
+   lifeTable$cumProbOfSurv[i] <-
+     lifeTable$condProbOfSurv[i]*lifeTable$condProbOfSurv[i-1]}
> #cumProbOfEvent
> lifeTable$cumProbOfEvent <- 1-lifeTable$cumProbOfSurv
```