

講義4

地域がん登録データを用いる デモンストレーション

国立がん研究センター
がん対策情報センター がん統計研究部
松田 彩子

集計対象

解析対象1

解析対象2

解析対象1から「がん死亡情報からの遡り調査による登録」を除外



このデータ年は遡り調査未実施のため、
解析対象2のみ計測可能

罹患数・率の集計対象から除くもの

第3期モニタリング項目およびコード区分

項目番号	項目名	変数名	変数型	区分
2	多重がんの有無	seq_no	数値型1桁	1 多重がんなし、あるいは第1がん 2 第2がん以降 9 多重がんか否かの区別不詳、多重がんの順番不詳
4	生年月	birth_dt	数値型6桁	} 罹患年齢 <div style="border: 2px solid red; border-radius: 10px; padding: 5px; display: inline-block; margin-left: 10px;">年齢不詳および100歳以上の症例</div>
5	診断年月	diag_dt	数値型6桁	
8	組織コード	icdom	数値型6桁	ICD-O-3の組織コードの6桁 不詳は、該当するコードを付与し(800039、801029など)、原則として空白は認めない。 <div style="border: 2px solid red; border-radius: 10px; padding: 5px; display: inline-block; margin-left: 10px;">5桁目が2</div>
11	DCO区分	dco_j	数値型1桁	<div style="border: 2px solid red; border-radius: 10px; padding: 5px; display: inline-block; margin-left: 10px;">1 DCO (死亡票のみ)</div> 2 DCO 以外
12	臨床進行度	extent	数値型1桁	<div style="border: 2px solid red; border-radius: 10px; padding: 5px; display: inline-block; margin-left: 10px;">1 上皮内がん</div> 2 限局 3 領域(所属リンパ節転移、隣接臓器浸潤) 4 遠隔転移 8 再発、DCO 9 不明、情報収集なし

参考

罹患集計用年齢計算式

作業用年齢 = [(診断年 × 12 + 診断月) - (生年 × 12 + 生月)] / 12
(診断年: 診断年月YYYYMMのYYYY, 生年: 生年月YYYYMMのYY, 生月: 生年月YYYYMMのMM)

- 診断月、または、生月が不明(99)のとき、作業用年齢=診断年-生年とする。
- 生年が不明(9999)の場合、作業用年齢計算はせず、年齢不詳とする。
- 小数点以下は切り捨てにする。

3年相対生存率を求める

$$\text{相対生存率} = \frac{\text{実測生存率}}{\text{期待生存率}}$$

EXCELの場合

- ① 診断日を起点とした実測生存率を、Kaplan-Meier法を用いて計算する。

使用するがん登録データ

実測 生存率	項目名	新規変数	使用する変数名
	診断月	mmdx	診断年月 : diag_dt
	診断年	yydx	
	死亡月	mmdeath	死亡年月 : death_dt
	死亡年	yydeath	
	生月	mmbirth	生年月 : birth_dt
	生年	yybirth	
	最終生存確認月	mmlife	最終生存確認年月 : life_dt
	最終生存確認年	yylife	
	最終状態月	mmexit	死亡年月 : death_dt
	最終状態年	yyexit	最終生存確認年月 : life_dt
	生存月数	surv_mm	最終状態年月(yyexit, yydx) 診断年月(mmexit, mmdx)
	生死状況 (死亡:1, 生存:0)	status	死亡年月 : death_dt 最終生存確認年月 : life_dt

● 最終生存確認年月

- 住民票照会以外で生存確認調査を実施している地域
 - 各登録より全死亡との照合対象年月として報告された日付を一律代入して利用

- 住民票照会を実施しない場合
 - 死亡情報がなかった症例は、全員3年生存とみなされる

- 生存確認調査を実施している住民票照会を実施している地域
 - 住民票照会の結果に基づく日付を利用

「診断からの生存月数」および 「生死状況」の作成

<p>生存月数</p>	<p>surv_mm</p>	<p>最終状態年月日作成</p> <p>死亡年月 (death_dt) を代入 死亡年月 (death_dt) が空欄の場合は2010年12月を代入 *** ****</p> <p>診断からの生存月数の作成</p> <p>(最終状態年 - 診断年) × 12 + (最終状態月 - 診断月)</p>
<p>生死状況 (死亡:1, 生存:0)</p>	<p>status</p>	<p>死亡年月空欄 → 0 死亡年月空欄以外 → 1</p>

②期待生存率は、0.5歳分加算したcohort生存率表に基づき、Ederer II法を用いて計算する。

地域がん登録集計用コホート生存率表

地域がん登録の技術支援のページ (http://ncrp.ncc.go.jp/pop_cohort_table.html)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	dy	age	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11
2	1	1962	0	0.98381	0.997743	0.998526	0.998831	0.999039	0.999243	0.99935	0.999405	0.999585	0.999677	0.9997
3	1	1962	1	0.9974	0.998346	0.99879	0.998949	0.999134	0.99934	0.999435	0.999506	0.999642	0.999712	0.9996
4	1	1962	2	0.998215	0.998613	0.998952	0.999106	0.999241	0.999427	0.999522	0.999582	0.999657	0.999703	0.9996
5	1	1962	3	0.998475	0.998823	0.999057	0.999252	0.999352	0.999502	0.999588	0.999618	0.999643	0.999643	0.9995
6	1	1962	4	0.998695	0.999004	0.999178	0.999368	0.999453	0.999553	0.999618	0.999618	0.999583	0.999537	0.9993
7	1	1962	5	0.998945	0.999164	0.999294	0.999459	0.999529	0.999578	0.999623	0.999578	0.999483	0.999306	0.9991
8	1	1962	6	0.999145	0.999294	0.999399	0.999524	0.999569	0.999574	0.999599	0.999498	0.999297	0.999055	0.9989
9	1	1962	7	0.999295	0.999395	0.999479	0.999559	0.999579	0.999534	0.999539	0.999328	0.999101	0.998909	0.9988
10	1	1962	8	0.999405	0.999495	0.999524	0.999559	0.999554	0.999474	0.999403	0.999142	0.998975	0.998798	0.9988
11	1	1962	9	0.999485	0.99955	0.99954	0.999529	0.999484	0.999363	0.999248	0.999016	0.998875	0.998722	0.9987
12	1	1962	10	0.99952	0.999535	0.99952	0.999474	0.999349	0.999238	0.999127	0.998905	0.998799	0.998727	0.9987
13	1	1962	11	0.9995	0.999495	0.999459	0.999374	0.999198	0.999132	0.999011	0.998814	0.998773	0.998751	0.9987
14	1	1962	12	0.999445	0.999425	0.999339	0.999254	0.999078	0.999012	0.998905	0.998753	0.998767	0.998745	0.9987
15	1	1962	13	0.99936	0.999325	0.999194	0.999138	0.998962	0.998896	0.998824	0.998707	0.998735	0.998744	0.9987
16	1	1962	14	0.999255	0.999209	0.999064	0.999028	0.998841	0.998789	0.998748	0.998661	0.998709	0.998738	0.9986
17	1	1962	15	0.99913	0.999084	0.998938	0.998917	0.998725	0.998688	0.998676	0.998624	0.998678	0.998727	0.9986
18	1	1962	16	0.998985	0.998949	0.998813	0.998791	0.998634	0.998602	0.99863	0.998598	0.998631	0.998705	0.9986
19	1	1962	17	0.998835	0.998804	0.998682	0.998867	0.998558	0.998541	0.998589	0.998582	0.998595	0.998684	0.9986

- ・sex ...性別(1: 男性, 2: 女性)
- ・dy ... 暦年(1962年～2009年)
- ・age ...年齢(0歳～99歳)[※0.5歳加算済み。例: age=0の場合0.5歳を意味する]
- ・s1 ...各行の性別・暦年・年齢の1年目(1年後までの1年間)の生存確率
- ・s2 ... 各行の性別・暦年・年齢の2年目(1年後から2年後までの1年間)の生存確率
- ・s3 ... 各行の性別・暦年・年齢の2年目(2年後から3年後までの1年間)の生存確率

使用するデータセット作成:

(期待生存率を求めるデータシートとコホート生命表をマージさせる)

期待 生存率	性別	sex	性別と年と罹患年を連結 入力例: E2&"-"&B2&"-"&F2 出力例: 1-2007-76
	年	year	
	罹患年齢	age	
	期待生存率用新規変数 (「コホート生命表」と「期待生存率を 求めるデータシート」、それぞれ作成)		

左端の列(**期待生存率用新規変数**)を検索して、「コホート生命表」の該当する列を「期待生存率を求めるデータシート」に貼り付ける。→ **VLOOKUP**を使用する

VLOOKUP(検索値, 範囲, 列番号, 検索の型)

検索値: 探したい値

範囲: 探したい範囲

列番号: 該当する列の位置

検索の型: テーブルを探す探し方; 完全一致→FALSE, 近似→TRUE

例: VLOOKUP(B2,生命表!\$A\$2:\$G\$9801,5,FALSE)

Rの場合

(別添付：Rプログラム)

使用するデータセット作成

まずデータの準備としてマージする両方のCSVでデータセット
(`R_2007_survival.csv`と`R_cohort_table_StdDBS_2010.csv`)において、
マージの際に使用する変数の名前を統一する("sex","year","age")。

`read.csv`という関数を使って`dat1`に`R_cohort_table_StdDBS_2010.csv`のなかのデータを読み込む。

```
dat1 = read.csv(" I:/DATA/ R_cohort_table_StdDBS_2010.csv", header=T)
```

データセットの
保存場所

データセット名

`read.csv`という関数を使って、`dat2`に`R_survival.csv`のなかのデータを読み込む。

```
dat2 = read.csv(" I:/DATA/ R_2007_survival.csv", header=T)
```

データセットの
保存場所

データセット名

#読み込んだRのデータセットである`dat1`(内容: `R_cohort_table_StdDBS_2010.csv`)と`dat2`(内
容: `R_2007_survival.csv`)をマージする。それを`dat3`に代入する。

```
dat3 = merge(dat1,dat2,by = c("sex", "year", "age"))
```

"sex","year","age"は最初に統一しておいたマージキー。

3年相対生存率を求める

#実測生存率をもとめる#

```
library(survival)
```

```
## Kaplan-Meier法による推定##
```

```
Surv(dat3$urv_mm, dat3$status)
```

```
alld.kap = survfit(Surv(surv_mm, status == 1) ~ 1, data = dat3, conf.int = 0.95)
```

```
alld.kap
```

```
summary(alld.kap)
```

→3年(36か月)生存率をもとめるので、timeが35のときのsurvivalの値をよみとる。

```
## 参考: 生存時間曲線##
```

```
plot(alld.kap, conf.int = F, main = "Kaplan-Meier (All case)",  
      xlab = "Survival Time", ylab = "Survival Probability")
```

#期待生存率をもとめる: 生命表#

```
## 1年生存率: 対象者; 2607名 ##
```

```
m_s1 = mean(dat3$s1)
```

```
## 2年生存率: 1年後の対象者; 2551名 ##
```

```
m_s2 = mean(dat3$s2[dat3$urv_mm >= 12])
```

```
## 3年生存率: 2年後の対象者; 2312名 ##
```

```
m_s3 = mean(dat3$s3[dat3$urv_mm >= 24])
```

```
### 3年期待生存率: 1年生存率 × 2年生存率 × 3年生存率###
```

```
m_s1 * m_s2 * m_s3
```

#3年相対生存率#

```
### 実測生存率 / 期待生存率: それぞれ求めた実測生存率と期待生存率を使用###
```

出力1

```
library(survival)
```

```
要求されたパッケージ splines をロード中です
```

```
Surv(dat3$surv_mm,dat3$status)
```

```
[省略]
```

```
alld.kap=survfit(Surv(surv_mm,status==1)~1,data=dat3,conf.int=0.95)
```

```
alld.kap
```

```
Call: survfit(formula = Surv(surv_mm, status == 1) ~ 1, data = dat3, conf.int = 0.95)
```

```
records n.max n.start events median 0.95LCL 0.95UCL
```

```
2607 2607 2607 880 49 49 51
```

```
summary(alld.kap)
```

```
Call: survfit(formula = Surv(surv_mm, status == 1) ~ 1, data = dat3, conf.int = 0.95)
```

time	n.risk	n.event	累積生存率 survival	標準誤差 std.err	下限信頼区間 lower 95% CI	上限信頼区間 upper 95% CI
------	--------	---------	-------------------	-----------------	------------------------	------------------------

```
[省略]
```

35	2023	26	0.76601	0.008292	0.749934	0.7824
----	------	----	----------------	-----------------	-----------------	---------------

36	1997	26	0.75604	0.008411	0.739734	0.7727
----	------	----	---------	----------	----------	--------

```
[省略]
```

出力2

#期待生存率をもとめる#

```
## 1年生存率 ##
```

```
m_s1=mean(dat3$s1)
```

```
m_s1
```

```
[1] 0.9718275
```

```
## 2年生存率:1年後の対象者 ##
```

```
m_s2=mean(dat3$s2[dat3$urv_mm >=12])
```

```
m_s2
```

```
[1] 0.9692034
```

```
## 3年生存率:2年後の対象者 ##
```

```
m_s3=mean(dat3$s3[dat3$urv_mm >=24])
```

```
m_s3
```

```
[1] 0.9680837
```

```
## 3年期待生存率:1年生存率 × 2年生存率 × 3年生存率##
```

```
m_s1*m_s2*m_s3
```

```
[1] 0.9118367
```

#3年相対生存率#

```
#### 実測生存率 / 期待生存率:それぞれ求めた実測生存率と期待生存率を使用####
```

```
0.76601 / 0.9118367
```

```
[1] 0.8400737
```

3年相対生存率74.5%

参考

library(survival)

##例:性別で群分けをする##

Kaplan-Meier 曲線をSEXで群分けする場合にはモデル式にsexを指定する。

ただし、sexが数値として読み込まれている場合は、FACTORとして変換する。

→新しい変数sex_codeを作成

```
dat3$sex_code=factor(dat3$sex)
kap_sex = survfit(Surv(surv_mm,status==1)~sex_code, data = dat3,conf.int=0.95)
kap_sex
summary(kap_sex)
plot(kap_sex,lty=1:2,col=2:3,conf.int=F,main="Kaplan-Meier estimate by SEX",
      xlab="Survival Time",ylab="Survival Probability")
```

#期待生存率をもとめる:生命表#

1年生存率 : 男女

```
m_s11=mean(dat3$s1[dat3$sex==1])
```

```
m_s12=mean(dat3$s1[dat3$sex==2])
```

2年生存率:1年後の対象者; 男女

```
m_s21=mean(dat3$s2[dat3$surv_mm >=12 & dat3$sex ==1])
```

```
m_s22=mean(dat3$s2[dat3$surv_mm >=12 & dat3$sex ==2])
```

3年生存率:2年後の対象者; 男女

```
m_s31=mean(dat3$s3[dat3$surv_mm >=24 & dat3$sex ==1])
```

```
m_s32=mean(dat3$s3[dat3$surv_mm >=24 & dat3$sex ==2])
```

3年期待生存率:1年生存率 × 2年生存率 × 3年生存率##

#男性#

```
m_s11*m_s21*m_s31
```

#女性#

```
m_s12*m_s22*m_s32
```