**2023 NTHU Data Science for Digital Health Homework 1**

**(Deadline: Nov. 3 23:59pm)**

#1 Predicting Mortality of Pulmonary Embolism Patients in the ICU

1. 18%

Adults (age ≥ 18) in ICU

with pulmonary embolism

(N = 2394)

Not die in hospital

(N = 2134)

Die in hospital

(N = 260)

1. 23%

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Outcome** | |
| Alive | Die |
| Gender | Male | 1070 | 111 |
| Female | 1064 | 149 |
| Age | 0-20 | 9 | 0 |
| 21-40 | 230 | 16 |
| 41-60 | 694 | 59 |
| 61-80 | 939 | 129 |
| 80+ | 262 | 56 |
| Race | Black | 306 | 38 |
| White | 1486 | 143 |
| Asian | 43 | 10 |
| Hispanic | 74 | 6 |
| Others | 225 | 63 |

如果對於一個病人有不同的種族，假設這個病人的不同種族中有這四種’BLACK’, ‘WHITE’, ‘ASIAN’, ‘HISPANIC’的其中一個的話，那我就選讓這個病人的種族是表格中病人被標示的”這四種其中之一的那一個”。例如：subject\_id=10036086的病人有兩個種族’WHITE’, ’PORTUGUESE’，那我就把這個病人的種族當作是’WHITE’。

Race的Alive的Others的算法是2134-306-1486-43-74 = 225。

#2 Renal Recovery Prediction Following Initiation of Renal Replacement Therapy

(1) 36%

Adults (age ≥ 18) in ICU and requiring dialysis (N = 1700)

dialysis\_type is ‘Peritoneal’

In hospital death

Included patients (N = 470)

Underlying CKD V or ESRD

(N = 1619)

(N = 1157)

Recover (N = 310)

Not recover (N = 160)

使用WHERE ordercategoryname LIKE '%Dialysis%'時，ordercategoryname只有兩種，一個是Dialysis，一個是Peritoneal Dialysis，所以要排除Peritoneal Dialysis時，我直接找ordercategoryname = Dialysis。

有160個人的order\_type有hemodialysis且icu\_outtime-last dialysis order(hemodialysis) time<72，有203人的order\_type有hemodialysis，有 470-203=267個人的order\_type沒有hemodialysis，所以icu\_outtime-last dialysis order(hemodialysis) time>=72hours的人數是(203-160)+267 = 310個人。

Note: 在找”所有的dialysis，包含hemodialysis”、age、gender、race時，需要把一些註解移走或加入。例如：在找”所有的dialysis，包含hemodialysis”時，就要把procedureevents\_dialysis相關的都註解。在判斷是否recover前，都不要有poe的出現。

Code以hw1\_1.txt和hw1\_2.txt為主，不過本檔案中的Code應該與hw1\_1.txt和hw1\_2.txt中的內容一樣。

(2) 23%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Outcome** | |  |
| Not recover  (<72hours) | Recover  (>=72hours or no hemodialysisorder\_type) | Total  470 |
| Gender | Male | 99 | 292-99 = 193 | 292 |
| Female | 61 | 178-61 = 117 | 178 |
| Age | 0-20 | 1 | 3-1 = 2 | 3 |
| 21-40 | 19 | 53-19 = 34 | 53 |
| 41-60 | 59 | 176-59 = 117 | 176 |
| 61-80 | 70 | 204-70 = 134 | 204 |
| 80+ | 11 | 34-11 = 23 | 34 |
| Race | Black | 16 | 48-16 = 32 | 48 |
| White | 95 | 308-95 = 213 | 308 |
| Asian | 6 | 13-6 = 7 | 13 |
| Hispanic | 8 | 19-8 = 11 | 19 |
| Others | 160-16-95-6-8 = 35 | 310-32-213-7-11 = 47 |  |

-- hw1\_1  --done10/29

SELECT COUNT(DISTINCT icd.subject\_id)

FROM

`physionet-data.mimiciv\_hosp.diagnoses\_icd` AS icd

INNER JOIN

(SELECT subject\_id, min(hospital\_expire\_flag) AS min\_hosp\_die\_flag -- min\_hosp\_die\_flag means first time left icu is dead or alive

FROM `physionet-data.mimiciv\_derived.icustay\_detail` GROUP BY subject\_id ) AS icu

ON icd.subject\_id = icu.subject\_id

INNER JOIN `physionet-data.mimiciv\_hosp.patients` AS age

ON icd.subject\_id = age.subject\_id

INNER JOIN `physionet-data.mimiciv\_derived.icustay\_detail` AS icu2 ON icd.subject\_id = icu2.subject\_id

WHERE (icd\_code LIKE 'I26%' OR icd\_code LIKE '4151%') AND (anchor\_age >= 18)

-- ans 2394

AND (icu.min\_hosp\_die\_flag = 0) -- 1:die in hospital, 0:not die in hospital

-- ans die in hospital=260

--AND age.gender = 'F' -- 'M':Male, 'F':Female

--AND age.anchor\_age BETWEEN 0 AND 20  -- age

AND race LIKE 'WHITE%'-- OR race LIKE 'BLACK%' OR race LIKE 'HISPANIC%' OR race LIKE 'ASIAN%'  -- race

-- hw1\_2 -- 11/3/11/22

SELECT COUNT(DISTINCT dialysis\_age18.subject\_id)

--SELECT DISTINCT dialysis\_age18.subject\_id, stay\_id, icu\_outtime, ordertime

--SELECT DISTINCT dislysis\_age18.subject\_id, ordercategoryname, icd.icd\_code, find\_dialysis.starttime, icd.icd\_version, age.anchor\_age

--select distinct subject\_id

FROM

  (SELECT icd.subject\_id, icd.icd\_code, age.anchor\_age, icu\_detail\_dialysis\_first.hospital\_expire\_flag, icu\_detail\_dialysis\_first.stay\_id, icu\_detail\_dialysis\_first.icu\_outtime,icu\_detail\_dialysis\_first.gender, icu\_detail\_dialysis\_first.race , ordertype\_Hemodialysis.ordertime,

  --ordertype\_Hemodialysis

  FROM

 `physionet-data.mimiciv\_hosp.diagnoses\_icd` AS icd

  INNER JOIN `physionet-data.mimiciv\_derived.age` AS age

  ON icd.subject\_id = age.subject\_id

  INNER JOIN

    (SELECT subject\_id, min(starttime) AS min\_starttime FROM

      (SELECT \* FROM `physionet-data.mimiciv\_icu.procedureevents` WHERE

      ordercategoryname LIKE '%Dialysis%')

    GROUP BY subject\_id) AS find\_min\_starttime

  ON icd.subject\_id = find\_min\_starttime.subject\_id

-- min\_starttime : the min starttime when dialysis or peritoneal dialysis happen

-- starttime means the event starts

  INNER JOIN `physionet-data.mimiciv\_icu.procedureevents` AS find\_dialysis

  ON find\_min\_starttime.subject\_id = find\_dialysis.subject\_id AND find\_min\_starttime.min\_starttime = find\_dialysis.starttime

-- find\_dialysis contains the data with dialysis or peritoneal dialysis from procedureevents

  INNER JOIN

    (SELECT subject\_id, min(starttime) AS min\_starttime\_dialysis FROM (SELECT \* FROM`physionet-data.mimiciv\_icu.procedureevents` WHERE

      ordercategoryname = 'Dialysis')

    GROUP BY subject\_id) AS procedureevents\_dialysis

  ON icd.subject\_id = procedureevents\_dialysis.subject\_id

  INNER JOIN `physionet-data.mimiciv\_icu.procedureevents` AS procedureevents\_dialysis2 ON procedureevents\_dialysis.subject\_id = procedureevents\_dialysis2.subject\_id AND procedureevents\_dialysis.min\_starttime\_dialysis = procedureevents\_dialysis2.starttime

  -- "the patients with dialysis" first in icu -> data

  INNER JOIN `physionet-data.mimiciv\_derived.icustay\_detail` AS icu\_detail\_dialysis\_first ON procedureevents\_dialysis2.subject\_id = icu\_detail\_dialysis\_first.subject\_id AND procedureevents\_dialysis2.stay\_id = icu\_detail\_dialysis\_first.stay\_id

  -- stay\_id of "the patients with dialysis" first in icu

-- find first icu outtime end

INNER JOIN (

  SELECT \* FROM `physionet-data.mimiciv\_hosp.poe` AS poe  WHERE poe.order\_type = 'Hemodialysis'

)AS ordertype\_Hemodialysis

ON icu\_detail\_dialysis\_first.subject\_id = ordertype\_Hemodialysis.subject\_id

   AND icu\_detail\_dialysis\_first.hadm\_id = ordertype\_Hemodialysis.hadm\_id

   AND ordertype\_Hemodialysis.ordertime <= icu\_detail\_dialysis\_first.icu\_outtime

-- find all ordertime before first icu\_outtime for each patients with dialysis and getting into icu first time

  WHERE anchor\_age >= 18 AND find\_dialysis.ordercategoryname LIKE '%Dialysis%'

-- ans 1700

  AND find\_dialysis.ordercategoryname = 'Dialysis'

-- ans 1619

-- dialysis\_age18 means the patients who have dialysis and age>=18

) AS dialysis\_age18 WHERE dialysis\_age18.subject\_id NOT IN (SELECT DISTINCT subject\_id FROM `physionet-data.mimiciv\_hosp.diagnoses\_icd` AS icd WHERE (icd\_code = '40301' OR icd\_code = '40311' OR icd\_code = '40391' OR icd\_code = '40402' OR icd\_code = '40403' OR icd\_code = '40412' OR icd\_code = '40413' OR icd\_code = '40492' OR icd\_code = '40493' OR icd\_code = '5856' OR icd\_code = 'I120' OR icd\_code = 'I1311' OR icd\_code = 'I132' OR icd\_code = 'N186'))

-- ans 1157

AND (dialysis\_age18.hospital\_expire\_flag = 0) -- 0:not die in hospital

-- ans 470

-- already included patients

AND TIMESTAMP\_DIFF(icu\_outtime, ordertime, hour) < 72 -- <72 means not recover

-- ans 160

-- have smaller than 72hours must not recover  Otherwise, recover.

--dialysis\_age18.

--AND gender = 'F' -- 'M':Male, 'F':Female

--AND anchor\_age BETWEEN 0 AND 20  -- age

--AND race LIKE 'HISPANIC%' -- OR race LIKE 'ASIAN%' OR race LIKE 'WHITE%' OR race LIKE 'BLACK%'  -- race