Bostat 218 Problem Set 1

Due Feb 07 @ 11:59PM in PDF by email

AUTHOR
Li Zhang @ UCLA

PUBLISHED
February 5, 2025

1 Set up

```
library(DatabaseConnector)
absoluteFileName <- file.path(getwd(), "../data", "synthetic.duckdb")

connection <- connect(dbms = "duckdb", server = absoluteFileName)
```

2 OMOP CDM

1. John is an African American man born on August 4, 1974. Define an entry in the `PERSON` table that encodes this information.

```
sql <- "
INSERT INTO person (
  person_id, gender_concept_id, year_of_birth, month_of_birth, day_of_birth,
  birth_datetime, race_concept_id, ethnicity_concept_id, location_id,
  provider_id, care_site_id, person_source_value, gender_source_value,
  gender_source_concept_id, race_source_value, race_source_concept_id,
  ethnicity_source_value, ethnicity_source_concept_id
)
VALUES (
  2, 8507, 1974, 8, 4, '1974-08-04 00:00:00', 8516, 38003564,NULL, NULL, NULL,
  NULL, 'MALE', 0, 'African American', 0, 'Not Hispanic or Latino', 0
);"

executeSql(connection, sql)
```

```
|
|                                                                      |   0%
|
|======================================================================| 100%
```

2. John enrolled in his current insurance on January 1st, 2015. The data from his insurance database were extracted on July 1st, 2019. Define an entry in the `OBSERVATION_PERIOD` table that encodes this information.

```
        sql2 <- "
          INSERT INTO observation_period(
            observation_period_id, person_id, observation_period_start_date,
            observation_period_end_date, period_type_concept_id
          )
          VALUES(
            2, 2, '2015-01-01', '2019-07-01', 44814722
          );
        "
        executeSql(connection, sql2)
```

```
|
|                                                                      |   0%
|
|======================================================================| 100%
```

3. John was prescribed a 30-day supply of Ibuprofen 200 MG Oral tablets (NDC code: 76168009520) on May 1st, 2019. Define an entry in the `DRUG_EXPOSURE` table that encodes this information.

```
        sql3 <- "
          INSERT INTO drug_exposure (
            drug_exposure_id, person_id, drug_concept_id,
            drug_exposure_start_date, drug_exposure_start_datetime,
            drug_exposure_end_date, drug_exposure_end_datetime,
            verbatim_end_date, drug_type_concept_id, stop_reason,
            refills, quantity, days_supply, sig,
            route_concept_id, lot_number, provider_id,
            visit_occurrence_id, visit_detail_id,
            drug_source_value, drug_source_concept_id, route_source_value
            )
          VALUES (
            1001, 2, 19078461,
            '2019-05-01', '2019-05-01 00:00:00',
            '2019-05-31', '2019-05-31 00:00:00',
            NULL, 38000177, NULL,
            NULL, NULL, 30, NULL,
            4132161, NULL, NULL,
            NULL, NULL,
            '76168009520', 583945, NULL
            );
        "
        executeSql(connection, sql3)
```

```
|
|                                                                      |   0%
|
|======================================================================| 100%
```

4. Using SQL and R, retrieve all records of the condition “Gastrointestinal hemorrhage” (with concept ID 192671) from the `Eunomia` dataset.

```
#set up
connectionDetails <- Eunomia::getEunomiaConnectionDetails()

library(DatabaseConnector)
connection_eu <- connect(connectionDetails)
```

```
sql4 <- "
  SELECT *
  FROM condition_occurrence
  WHERE condition_concept_id = 192671;
"
gastro_records <- renderTranslateQuerySql(connection_eu, sql4)

head(gastro_records)
```

```
  CONDITION_OCCURRENCE_ID PERSON_ID CONDITION_CONCEPT_ID CONDITION_START_DATE
1                    4657       273                192671           2011-10-10
2                    1021        61                192671           2005-09-15
3                    5978       351                192671           2018-06-28
4                    9798       579                192671           1999-11-06
5                    9301       549                192671           1987-12-28
6                    1997       116                192671           1970-03-12
  CONDITION_START_DATETIME CONDITION_END_DATE CONDITION_END_DATETIME
1              2011-10-10              <NA>                  <NA>
2              2005-09-15              <NA>                  <NA>
3              2018-06-28              <NA>                  <NA>
4              1999-11-06              <NA>                  <NA>
5              1987-12-28              <NA>                  <NA>
6              1970-03-12              <NA>                  <NA>
  CONDITION_TYPE_CONCEPT_ID CONDITION_STATUS_CONCEPT_ID STOP_REASON PROVIDER_ID
1                     32020                          0        <NA>          NA
2                     32020                          0        <NA>          NA
3                     32020                          0        <NA>          NA
4                     32020                          0        <NA>          NA
5                     32020                          0        <NA>          NA
6                     32020                          0        <NA>          NA
  VISIT_OCCURRENCE_ID VISIT_DETAIL_ID CONDITION_SOURCE_VALUE
1               18192               0                  K92.2
2                4183               0                  K92.2
3               23432               0                  K92.2
4               38298               0                  K92.2
5               36419               0                  K92.2
6                7655               0                  K92.2
  CONDITION_SOURCE_CONCEPT_ID CONDITION_STATUS_SOURCE_VALUE
1                    35208414                          <NA>
2                    35208414                          <NA>
```

```
3                    35208414                        <NA>
4                    35208414                        <NA>
5                    35208414                        <NA>
6                    35208414                        <NA>
```

5. Using SQL and R, retrieve all records of the condition “Gastrointestinal hemorrhage” using source codes. This database uses ICD-10, and the relevant ICD-10 code is “K92.2” from the `Eunomia` dataset.

```
sql5 <- "
  SELECT *
  FROM condition_occurrence
  WHERE condition_source_value = 'K92.2';
"
gastro_records_source <- renderTranslateQuerySql(connection_eu, sql5)
head(gastro_records_source)
```

```
  CONDITION_OCCURRENCE_ID PERSON_ID CONDITION_CONCEPT_ID CONDITION_START_DATE
1                    4657       273               192671           2011-10-10
2                    1021        61               192671           2005-09-15
3                    5978       351               192671           2018-06-28
4                    9798       579               192671           1999-11-06
5                    9301       549               192671           1987-12-28
6                    1997       116               192671           1970-03-12
  CONDITION_START_DATETIME CONDITION_END_DATE CONDITION_END_DATETIME
1              2011-10-10              <NA>                  <NA>
2              2005-09-15              <NA>                  <NA>
3              2018-06-28              <NA>                  <NA>
4              1999-11-06              <NA>                  <NA>
5              1987-12-28              <NA>                  <NA>
6              1970-03-12              <NA>                  <NA>
  CONDITION_TYPE_CONCEPT_ID CONDITION_STATUS_CONCEPT_ID STOP_REASON PROVIDER_ID
1                     32020                           0        <NA>          NA
2                     32020                           0        <NA>          NA
3                     32020                           0        <NA>          NA
4                     32020                           0        <NA>          NA
5                     32020                           0        <NA>          NA
6                     32020                           0        <NA>          NA
  VISIT_OCCURRENCE_ID VISIT_DETAIL_ID CONDITION_SOURCE_VALUE
1               18192               0                 K92.2
2                4183               0                 K92.2
3               23432               0                 K92.2
4               38298               0                 K92.2
5               36419               0                 K92.2
6                7655               0                 K92.2
  CONDITION_SOURCE_CONCEPT_ID CONDITION_STATUS_SOURCE_VALUE
1                    35208414                         <NA>
2                    35208414                         <NA>
3                    35208414                         <NA>
4                    35208414                         <NA>
```

```
5                   35208414                    <NA>
6                   35208414                    <NA>
```

6. Using SQL and R, retrieve the observation period of the person with `PERSON_ID` 61 from the `Eunomia dataset.

```
sql6 <- "
  SELECT *
  FROM @cdm.observation_period
  WHERE person_id = 61;
"

renderTranslateQuerySql(connection_eu, sql6, cdm = "main")
```

```
  OBSERVATION_PERIOD_ID PERSON_ID OBSERVATION_PERIOD_START_DATE
1                    61        61                     1968-01-21
  OBSERVATION_PERIOD_END_DATE PERIOD_TYPE_CONCEPT_ID
1                  2019-01-06                44814724
```

```
disconnect(connection_eu)
```

3 Standardize vocabularies

7. What is the standard concept ID for “Gastrointestinal hemorrhage”?

- 192671

8. Which ICD-10CM codes map to the standard concept for “Gastrointestinal hemorrhage”? Which ICD-9CM codes map to this Standard Concept?

ICD-10CM

- K29.91: Gastroduodenitis, unspecified, with bleeding
- K92.2: Gastrointestinal hemorrhage, unspecified

ICD-9CM

- 578: Gastrointestinal hemorrhage
- 578.9: Hemorrhage of gastrointestinal tract, unspecified

9. What are the MedDRA preferred terms that are equivalent to the standard concept for “Gastrointestinal hemorrhage”?

- “Gastrointestinal haemorrhage” (Concept ID 35707864)
- “Intestinal haemorrhage” (Concept ID 35707858)

4 Advanced SQL

10. What is the minimum, maximum, and mean length (in days) of observation from the `synthetic` dataset? (Hint: you can use the `DATEDIFF` function to compute the time between two dates.)

```
library(DatabaseConnector)
absoluteFileName <- file.path(getwd(), "../data", "synthetic.duckdb")

connection <- connect(dbms = "duckdb", server = absoluteFileName)

sql10 <- "
  SELECT
    MIN(
      DATEDIFF(DAY, observation_period_start_date, observation_period_end_date)
      ) AS min_observation_days,
    MAX(
      DATEDIFF(DAY, observation_period_start_date, observation_period_end_date)
      ) AS max_observation_days,
    AVG(
      DATEDIFF(DAY, observation_period_start_date, observation_period_end_date)
      ) AS mean_observation_days
  FROM observation_period;
"
renderTranslateQuerySql(connection, sql10)
```

```
  MIN_OBSERVATION_DAYS MAX_OBSERVATION_DAYS MEAN_OBSERVATION_DAYS
1                    0                40509              13683.64
```

11. How many people have at least one prescription of celecoxib from the `synthetic` dataset? (Note: there's an easy way to do this, using `DRUG_ERA`, and a harder way using `DRUG_EXPOSURE` and `CONCEPT_ANCESTOR`. Can you do both?)

```
sql11 <-"
  SELECT COUNT(DISTINCT(person_id)) AS num_people
  FROM drug_era
  WHERE drug_concept_id = 1118084
"
renderTranslateQuerySql(connection, sql11)
```

```
  NUM_PEOPLE
1          0
```

```
sql11_1 <- "
  SELECT COUNT(DISTINCT de.person_id) AS num_people
  FROM drug_exposure de
  JOIN concept_ancestor ca ON de.drug_concept_id = ca.descendant_concept_id
  WHERE ca.ancestor_concept_id = 1118084;
"
```

```
renderTranslateQuerySql(connection, sql11_1)
```

```
  NUM_PEOPLE
1          0
```

12. During which period in time (calender start and end date) did people start a celecoxib prescription from the `synthetic` dataset?

```
sql12 <- "
  SELECT
    MIN(drug_era_start_date) AS first_prescription_date,
    MAX(drug_era_start_date) AS last_prescription_date
  FROM drug_era
  WHERE drug_concept_id = 1118084
"

renderTranslateQuerySql(connection, sql12)
```

```
  FIRST_PRESCRIPTION_DATE LAST_PRESCRIPTION_DATE
1                    <NA>                   <NA>
```

```
disconnect(connection)
```