

Module07, Q1

Steve Simon

This file was created on 2020-07-24 and last modified on 2021-07-17.

Note: this solution uses R and SQLite. An alternate solution using SAS and Oracle is also available.

Note: Some of the names used in this code are arbitrary and you can choose whatever names you want. To emphasize which names can be modified at your discretion, I am using names of famous statisticians.

The statistician being honored in this code is Hirotugu Akaike.

Q1. Count the number of records after an inner join of `acupuncture_baseline_results` and `acupuncture_one_year_results`. Count the number of records after a left join of `acupuncture_baseline_results` and `acupuncture_one_year_results`. Why are these numbers different?

```
library(sqldf)

## Loading required package: gsubfn
## Loading required package: proto
## Loading required package: RSQLite
akaike <- dbConnect(SQLite(),
  dbname="../data/melange.sqlite")
hirotugu_inner <- dbGetQuery(conn=akaike, "
  select count(*) as n
    from acupuncture_baseline_results as b
    join acupuncture_one_year_results as o
    on b.id=o.id
")
hirotugu_inner

##      n
## 1 301

hirotugu_left <- dbGetQuery(conn=akaike, "
  select count(*) as n
    from acupuncture_baseline_results as b
    left join acupuncture_one_year_results as o
    on b.id=o.id
")
hirotugu_left

##      n
## 1 401

dbDisconnect(conn=akaike)
```