

# ACMA 231 – Exam 3

---

## Part 1: SQL (25 pts)

For this part, you will use a schema named 'Narnia' in the SQL\_Bpp that contains three tables. Note that **claims** and **payments** tables are linked with **RecNo** field, and **policies** and **claims** tables are linked with **PolicyNo** field.

1. Write a query that aggregates **Payment** with sum for each of the **RecNo** from the **payments** table. (5 pts)
2. Write a query that inner joins **claims** table and the query from Question 1 on **RecNo**. (5 pts)
  - The output will include two fields, **PolicyNo** and **Pmt** (renamed field of the aggregated payment from the query from Question 1).
3. Write a query that aggregates **Pmt** with sum for each of the **PolicyNo** from the query from Question 2. (5 pts)
  - The output will include two fields, **PolicyNo** and **annualPmt** (renamed field of the aggregated payment from the query from Question 2).
4. Write a query that inner joins **policies** table and the query from Question 3 on **PolicyNo**. (10 pts)
  - Only include policies where **annualPmt** is greater than 0.
  - Only include policies where **FleetFlag** is "N".
  - A new field, called "**VehicleType**" is created that will do the following:
    - If the **VehicleCode** is G or S, provide "Trucks"
    - If the **VehicleCode** is M, provide "Motorcycle"
    - Otherwise, provide "Cars"
  - A new field, called "**VehicleYear**" is created that provides the difference 1993 from **ManufactureYr** in the **policy** table.
  - Include the following fields in the output table:
    - The **PolicyNo**, **VehicleCapacity**, **AgeInsured**, **SexInsured**, **Marital**, and **NCD** fields directly from the **policy** table,
    - The **VehicleYear** and **VehicleType** fields derived from the **policy** table,
    - The **annualPmt** field summarized from the **claims** and **payment** tables.
  - Final output is given ascending alphabetically by **PolicyNo**.

## Part 2: R (25 pts)

For this part, you will use the extracted table (which can be downloaded as a csv file) from Question 4 of Part 1.

5. Load the csv file in R as a dataframe and remove any redundant variable(s) in the dataframe. (2 pts)
6. Add a covariate named  $\ln\text{VehicleCapa}$ , which is defined as  $\log(1 + \text{VehicleCapacity})$ . (3 pts)
7. Conduct an explanatory analysis to see whether there are noticeable differences among the distributions of  $\text{VehicleCapacity}$  per  $\text{VehicleType}$ . (5 pts)
8. Split the dataset into two parts; 80% of training set and 20% of test set using 'createDataPartition' function. When you split the data, please fix the random seed to 1000. (3 pts)
9. Fit the following regression models: (8 pts)
  - Model 1: Assume that  $\text{annualPmt}$  follows a gamma distribution with log link, and the explanatory variables (covariates) are NCD,  $\text{AgeInsured}$ ,  $\text{SexInsured}$ ,  $\ln\text{VehicleCapa}$ , and  $\text{VehicleType}$ . (no interaction terms among the covariates)
  - Model 2: Use  $\text{AgeInsured}^2$  on top of Model 1.
  - Model 3: Use the interaction term of  $\ln\text{VehicleCapa}$  and  $\text{VehicleType}$  on top of Model 2.
10. Compare the prediction performance of the three models in Question 9 using RMSE. (4 pts)