# ASSIGNMENT 3

## ISYS1055/1057 DATABASE CONCEPTS

Yogesh Haresh Bojja
S3789918 | 06-06-2020

- **DATA PREPARATION**

  - **Introduction –** We are given with the data that contains survey questions about the course evaluation, course details and staff details. We have 455 courses in our dataset. One academic can teach many courses. Academic having beauty score more than 0 is judged to be more attractive than the average person. In this assignment we are going to investigate whether features like class size, attractiveness of teaching staff etc influences course evaluation or not.

  - **Relation schema –** As there is data redundancy in staffid, age, tenure track, beauty, gender, and non-English columns we need to normalize or decompose the table. Our table is in 1NF form as all the values are atomic.
    **Functional dependencies in the table are:**
    - id → staffid, courseevaluation, students, division
    - staffid → age, gender, tenure track, beauty, nonenglish

    As there are no partial dependencies present we can say that the table is in 2NF form. We can see that there is transitive dependency present i.e. id → staffid and staffid → age, gender, tenure track, beauty, nonenglish. Hence we need to decompose the table to normalize it in 3NF form.

    After decomposition, the schema of tables are:

    - staffDetails (<u>staffid</u>, age, gender, tenuretrack, nonenglish, beauty)
    - courseDetails (<u>id</u>, staffid*, courseevaluation, students, division)

    As all the prime attributes derive non-prime attributes and they all are candidate key we can conclude that the above schema after decomposition is in BCNF.

    After decomposition sizes of our tables are:

    - profEval = 455 rows
    - staffDetails = 90 rows
    - courseDetails = 455 rows

  - **Import csv file –** Right click on subsection tables in the project created and select import data. Locate the path of csv file. Click next on data preview tab. In the import method tab give the desired table name in my case table name is profEval. In choose columns tab select all the columns you wish to import from the csv file. In column definition tab we get to select the datatype and its size for each column, for example I have selected integer for staffid. While decomposition we create two new tables – staffDetails and courseDetails – by CREATE AS statements which is shown in appendix.

- **ANALYSIS**

  - **Course Sizes – Number of Students**

| | **Minimum** | **Mean** | **Maximum** |
|---|---|---|---|
| Number of Students | 8 | 55.378 | 581 |

    Even though the mean count of all students enrolled in a course is 55.378 there is a larger gap between the minimum course size and maximum course size which tells us that students tend to like same subject hence the count of such subject is more.

  - **Course Sizes – Course Evaluation Score**

| **Course size** | **18 or less** | **19-28** | **29-60** | **61 or more** |
|---|---|---|---|---|
| Number of course in group | 112 | 114 | 116 | 113 |

| | | | | |
|---|---|---|---|---|
| Minimum course evaluation score | 2.3 | 2.7 | 2.1 | 2.8 |
| Mean course evaluation score | 4.154 | 4.001 | 3.937 | 3.905 |
| Maximum course evaluation score | 5 | 5 | 5 | 4.8 |

We can see that course evaluation score decreases as the course size increases. There are almost same number of courses present in each course size section.

- **Division**

| | No. courses in group | Minimum | Mean | Maximum |
|---|---|---|---|---|
| Upper division | 300 | 2.1 | 3.951 | 5 |
| Lower division | 155 | 2.5 | 4.092 | 5 |

Upper division has more number of courses as compared to Lower division. Students studying courses of Lower division tend to rate course evaluation rating high as its mean and minimum value is higher than that of upper division.

- **Gender – Course Evaluation Scores**

| | No. courses in group | Minimum | Mean | Maximum |
|---|---|---|---|---|
| Female | 193 | 2.3 | 3.897 | 4.9 |
| Male | 262 | 2.1 | 4.073 | 5 |

Number of courses taught by male staff is more than the female staff. Male staff seem to teach the course well than that of female staff as its mean of course evaluation is high.

- **Gender – Beauty**

| | No. academics in group | Minimum | Mean | Maximum |
|---|---|---|---|---|
| Female | 39 | -1.538843 | 0.125 | 1.8816743 |
| Male | 51 | -1.51126770 | -0.115 | 1.6859847 |

There are more male academics than the female academics. Female academics are significantly more beautiful than the male academics.

- **Tenure Track**

| | No. academics in group | Minimum | Mean | Maximum |
|---|---|---|---|---|
| Tenure track | 75 | 2.1 | 3.96 | 5 |
| Not tenure track | 15 | 2.8 | 4.133 | 5 |

Out of total 90 academics 75 of them are working towards a permanent position. Academics not working towards a permanent position seem to be more effective in teaching the course as their mean of course evaluation is higher than that of the academics working towards their permanent position.

- **Education Background**

| | No. academics in group | Minimum | Mean | Maximum |
|---|---|---|---|---|
| English education | 83 | 2.1 | 4.019 | 5 |
| Non – English education | 7 | 2.7 | 3.689 | 4.6 |

There are more academics who did their undergraduate education in English speaking nation. Academics who studied in English speaking nation tend to teach the course well even though their minimum rating for course evaluation is less.

- **Interaction between Tenure Track, Gender and Education Background**

| Tenure track | Gender | Education | No. academics in group | Mean |
|---|---|---|---|---|
| Tenure track | Female | English | 28 | 3.928 |
| Tenure track | Female | Non-English | 3 | 3.717 |
| Tenure track | Male | English | 40 | 4.02 |
| Tenure track | Male | Non-English | 4 | 3.669 |
| Not tenure track | Female | English | 8 | 3.86 |
| Not tenure track | Female | Non-English | 0 | 0 |
| Not tenure track | Male | English | 7 | 4.396 |
| Not tenure track | Male | Non-English | 0 | 0 |

There are no academics from Non-English-speaking country who are not working towards a permanent position. Most of the academics are working towards a permanent position, out of which most of them did their undergraduate studies in English speaking country. Academics who did their studies in English speaking country are more effective, of which male academics top the list.

- **Correlation Analysis**

| Variables | Correlation Coefficient | Two-sided Significance |
|---|---|---|
| Course evaluation score & course size | -0.17286 | 0.00021 |
| Staff age & beauty | -0.33099 | 0.00144 |
| Staff age and mean course evaluation | -0.02214 | 0.83594 |
| Staff beauty and mean course evaluation score | 0.17851 | 0.0923 |

As the course size increase course evaluation score tends to decrease. More is the age of academics less beauty score do they get. Third observation shows that as the age increases, less effective do the academic seem which cannot be trusted as its p-value is greater than 0.05 i.e. 0.83594 > 0.05. More attractive is the academic, higher course evaluation they get is observed from fourth observation but we can't rely on it as p-value is greater than 0.05.

- **DISCUSSION AND CONCLUSION**

  – One part of our investigation was whether course size influence course evaluation. Observing the correlation analysis of these two variables we can say that there is inverse relationship between these two variables. As the course size increases by one-unit course evaluation decreases by 0.173. We can concur with this analysis as p-value of two-tailed test is 0.00021 which is less than the area of significance i.e. 0.05. Second part of our investigation was whether teaching evaluations are influenced by attractiveness of the teaching staff. We can see positive relationship between both these variables i.e. as beauty increases by one-unit course evaluation increases by 0.1785, but we cannot trust this correlation as p-value for its two-tailed test is 0.0923 which is greater than 0.05. Hence by studying the above data we can conclude that as the course size increases course evaluation decreases, but we don't have enough evidences to conclude that there is a significant relationship between attractiveness of the academic and course evaluation score.

- Limitations: As provided data is a subset of the original data collected our analysis may vary and influence to show different results as compared to the original results. We need to confirm our results for other analysis too by hypothesis testing with correlation & p-value. As our goal was to investigate the relationship between course size, attractiveness of the academics with the course evaluation score, I feel that analysis of course evaluation with other columns is not necessary.

**APPENDIX**

- **DATA PREPARATION**
  - Import data from csv file into prof Eval table from SQL developers GUI.
  - Decompose profEval into staffDetails and courseDetails and add constraints:
    - ```
      CREATE TABLE staffDetails
      AS
      SELECT staffid, age, gender, tenuretrack, nonenglish, beauty
      FROM profEval
      GROUP BY staffid, age, gender, tenuretrack, nonenglish, beauty;
      ```
    - ```
      ALTER TABLE staffDetails
      ADD CONSTRAINT sid_pk PRIMARY KEY (staffid);
      ```
    - ```
      CREATE TABLE courseDetails
      AS
      SELECT id, staffid, courseevaluation, students, division
      FROM profEval
      GROUP BY id, staffid, courseevaluation, students, division;
      ```
    - ```
      ALTER TABLE courseDetails
      ADD CONSTRAINT cid_pk PRIMARY KEY (id);
      ```
    - ```
      ALTER TABLE courseDetails
      ADD CONSTRAINT sid_fk FOREIGN KEY (staffid) REFERENCES staffDetails(staffid);
      ```
  - Check whether data is lossless by performing join on two tables and comparing it with original table:
    - ```
      SELECT id, s.staffid, age, gender, tenuretrack, nonenglish, courseevaluation, students,
      division, beauty
      FROM staffDetails s, courseDetails c
      WHERE s.staffid = c.staffid
      ORDER BY 1;
      ```
  - Check size and schema of tables:
    - `SELECT COUNT(*) AS "Num of rows" FROM profeval;`
    - `SELECT COUNT(*) AS "Num of rows" FROM staffdetails;`
    - `SELECT COUNT(*) AS "Num of rows" FROM coursedetails;`
    - `describe staffDetails;`
    - `describe courseDetails;`
    - `describe profEval;`

- **ANALYSIS**
  - Course sizes – Number of Students
    - ```
      SELECT MIN(students), ROUND(AVG(students), 3), MAX(students)
      FROM courseDetails;
      ```
  - Course sizes – Course Evaluation Score
    - ```
      SELECT COUNT(ID), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3),
      MAX(courseevaluation)
      FROM staffDetails s, courseDetails c
      WHERE s.staffid = c.staffid
      AND students <= 18;
      ```

- SELECT COUNT(ID), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
  FROM staffDetails s, courseDetails c
  WHERE s.staffid = c.staffid
  AND students > 18 AND students < 29;

- SELECT COUNT(ID), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
  FROM staffDetails s, courseDetails c
  WHERE s.staffid = c.staffid
  AND students > 28 AND students < 61;

- SELECT COUNT(ID), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
  FROM staffDetails s, courseDetails c
  WHERE s.staffid = c.staffid
  AND students > 60;

- Division
  - SELECT division, COUNT(ID), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
    FROM courseDetails
    GROUP BY division;

- Gender – Course Evaluation Score
  - SELECT gender, COUNT(ID), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
    FROM staffdetails s, courseDetails c
    WHERE s.staffid = c.staffid
    GROUP BY gender;

- Gender – Beauty
  - SELECT gender, COUNT(staffID), MIN(beauty), ROUND(AVG(beauty), 3), MAX(beauty)
    FROM staffdetails
    GROUP BY gender;

- Tenure Track
  - SELECT tenuretrack, COUNT(distinct(s.staffid)), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
    FROM staffdetails s, courseDetails c
    WHERE s.staffid = c.staffid
    GROUP BY tenuretrack;

- Education Background
  - SELECT nonenglish, COUNT(distinct(s.staffid)), MIN(courseevaluation), ROUND(AVG(courseevaluation), 3), MAX(courseevaluation)
    FROM staffdetails s, courseDetails c
    WHERE s.staffid = c.staffid
    GROUP BY nonenglish;

- Interaction between Tenure Track, Gender and Education Background
  - SELECT tenuretrack AS "Tenure Track", gender AS "Gender", nonenglish AS "Education", COUNT(distinct(s.staffid)) AS "No of academics in group", ROUND(AVG(courseevaluation), 3) AS "Mean"
    FROM staffdetails s, courseDetails c
    WHERE s.staffid = c.staffid
    GROUP BY tenuretrack, gender, nonenglish
    ORDER BY 1;

- Correlation Analysis
  - SELECT ROUND(CORR_S(courseevaluation, students, 'COEFFICIENT'), 5) AS "Coefficient", ROUND(CORR_S(courseevaluation, students, 'TWO_SIDED_SIG'), 5) AS "Two sided significance"
    FROM courseDetails;

  **Note: Create View staffView for further three correlation analysis**

  - CREATE OR REPLACE VIEW staffView (staffID, age, beauty, courseevaluation)
    AS
    SELECT s.staffid, age, beauty, ROUND(AVG(courseevaluation), 5)
    FROM staffdetails s, courseDetails c
    WHERE s.staffid = c.staffid
    GROUP BY s.staffid, age, beauty;

  - SELECT ROUND(CORR_S(age, beauty, 'COEFFICIENT'), 5) AS "Coefficient",
    ROUND(CORR_S(age, beauty, 'TWO_SIDED_SIG'), 5) AS "Two sided significance"
    FROM staffView;

  - SELECT ROUND(CORR_S(age, courseevaluation, 'COEFFICIENT'), 5) AS "Coefficient",
    ROUND(CORR_S(age, courseevaluation, 'TWO_SIDED_SIG'), 5) AS "Two sided significance"
    FROM staffView;

  - SELECT ROUND(CORR_S(beauty, courseevaluation, 'COEFFICIENT'), 5) AS "Coefficient",
    ROUND(CORR_S(beauty, courseevaluation, 'TWO_SIDED_SIG'), 5) AS "Two sided significance"
    FROM staffView;

**REFERENCES**


- [1] ThatJeffSmith(2012)."How to Import from Excel to Oracle with SQL Developer" – https://www.thatjeffsmith.com/archive/2012/04/how-to-import-from-excel-to-oracle-with-sqldeveloper/comment-page-5/ (Accessed: June 9 2020).
- [2]"CORR_*" https://docs.oracle.com/cd/B28359_01/server.111/b28286/functions029.htm#SQLRF51302 (Accessed: June 9 2020).