MSSE **Data** C**200**, Spring **202**1 February 8**, 202**1

Discussion #2

Name:

Guessing at Random

> A multiple choice test has 100 questions, each with five possible answers of which one is right. The grading scheme is as follows:4 points are awarded for each right answer. For each other answer (wrong, missing, etc), one point is taken off; that is, -1 points are awarded.

A student hasn't studied at all and therefore guesses each answer uniformly at random, independently of all the other answers.

Define the following random variables:

- R: the number of answers the student gets right
- W: the number of answers the student does not get right
- S: the student's score on the test

1. What is the distribution of R? Either state the possible values and provide a formula for the probabilities, or provide the name and parameters of the appropriate distribution. Explain your answer.
2. Find $\mathbb{E}(R)$.
3. Find $\mathbb{E}(S)$.

SQL Primary and Foreign Key

Definitions:

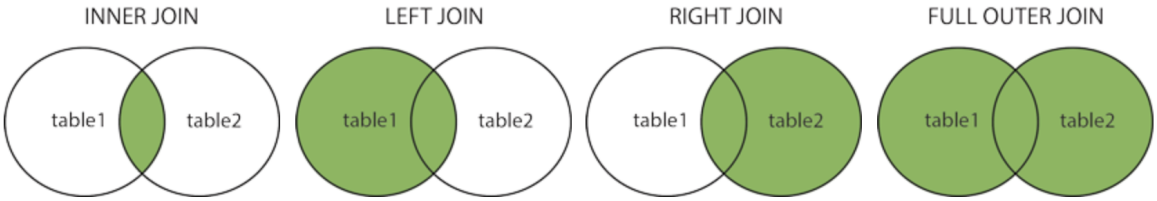
- **Primary key**: The column or minimal set of columns that uniquely determines the values in all the remaining columns. This is a statement about the schema and should hold for all data that could be put in the table.
  Below are some constraints on the primary key:
  - The data within these columns must be unique.
  - No value in the columns can be NULL.
- **Foreign key**: A set of one or more columns in a table that refers to the primary key in another table.
  Foreign keys have the following properties:
  - We can have NULL values in foreign keys.
  - We can have non-unique foreign keys in a table.
  - If the foreign key is not NULL, it should reference a particular primary key in another table.

4. **Examples?** What might be good examples of common primary keys and how might they be referenced as foreign keys.

5. Consider the following *sample* of the baby names table. What is the primary key? (Note: Assume the full table has more than one state, more than one sex, and more than one year)

	State	Sex	Year	Name	Count
0	CA	F	1910	Mary	295
1	CA	F	1910	Helen	239
2	CA	F	1910	Dorothy	220
3	CA	F	1910	Margaret	163
4	CA	F	1910	Frances	134

SQL Joins



Note: You do not always have to use the JOIN keyword to join sql tables. The following are equivalent:

```
SELECT column1, column2
FROM table1, table2
WHERE table1.id = table2.id;

SELECT column1, column2
FROM table1 JOIN table2
ON table1.id = table2.id;
```

6. Describe which records are returned from each type of join in the figure above. How does a cross join relate to these types of joins?

7. Consider the following real estate schema:

```
Homes(home_id int, city text, bedrooms int, bathrooms int,
area int)
Transactions(home_id int, buyer_id int, seller_id int,
transaction_date date, sale_price int)
Buyers(buyer_id int, name text)
Sellers(seller_id int, name text)
```

Fill in the blanks in the SQL query to find the id and selling price for each home in Berkeley. If the home has not been sold yet, **the price should be NULL**.

```
SELECT _____
FROM _____
_____ JOIN _____
ON _____
WHERE _____;
```

SQL Queries

8. Examine this schema for these two tables:

```
CREATE TABLE owners (
    ownerid integer,
    name text,
    age integer,
    PRIMARY KEY (ownerid)
);
```

```
CREATE TABLE cats (
    catid integer,
    owner integer,
    name text,
     breed text,
    age integer,
    PRIMARY KEY (catid),
    FOREIGN KEY (owner) REFERENCES owners
);
```

   (a) Write a SQL query to figure out the number of cats, over the age of 10, of each breed of cat.

   (b) Write a SQL query to figure out the number of cats each owner owns for owners whose id is greater than 10.

   (c) Write a SQL query to figure out the ownerid/owner of the one cat owner who owns the most cats.

   (d) Write a SQL query to figure out the names of all of the cat owners who have a cat named Apricot.

   (e) It is possible to have a cat with an owner not in the owners table.

   ◯ A. True ◯ B. False

(f) Write a SQL query to get a random sample of 5 random Maine Coons (a cat breed) with a name that starts with the letter A.

(g) (Challenge) Write a SQL query to create an almost identical table as cats, except with an additional column 'Nickname' that has the value 'Kitten' for cats less than or equal to the age of 1, 'Catto' for cats between 1 and 15, and 'Wise One' for cats older than or equal to 15.

(h) (Challenge) Write a SQL query to select all rows from the `cats` table that have cats of the top 5 most popular cat breeds.