Question 1. Short Questions (15 points)

1. Consider two relations given below. State if it is possible to perform the set union of these two relations i.e. R1 ∪ R2. If union is possible provide the resulting schema after union, if not, state your reasoning.

   R1: (FirstName (char), LastName(char), DatefBirth(date))
   R2: (FName(char), LName(char), PhoneNumber(number))

   Solution: It is not possible because R1 and R2 must have the same schema to perform union.

2. Use relational algebra to show 5 different ways to express: $\sigma_{c>20}\ (R \bowtie S)$. You can use basic and derived relational algebra operations covered in class.

   Solution: Lets assume the schema to be R(A,B) and S(B,C)

   (a) $R \bowtie (\sigma_{C>20}(S))$
   (b) $\sigma_{C>20}(\pi_{A,B,C}(\sigma_{B=B1}(R \times \rho_{B1,C}(S))))$
   (c) $\pi_{A,B,C}(\sigma_{B=B1}(R \times \rho_{B1,C}(\sigma_{C>20}(S))))$
   (d) $\sigma_{C>20}(\pi_{A,B,C}(R \bowtie_{B=B1} \rho_{B1,C}(S)))$
   (e) $\pi_{A,B,C}(R \bowtie_{B=B1} \rho_{B1,C}(\sigma_{C>20}(S)))$
   (f) $(R \bowtie S) - \sigma_{C\leq 20}(R \bowtie S)$

3. The following formulae are valid under set semantics. For each formula, mention if it is valid under bag semantics as well. If not, provide a counter example that demonstrates that it is not valid under bag semantics.

   (a) R ∪ S = S ∪ R
   (b) (R ∪ S) - T = (R - T) ∪ (S - T)
   (c) (R ∩ S) - T = R ∩ (S - T)

   Solution:(a) Yes (b) No. e.g. R = S = T = {x} (c) No. e.g. $R = T = \{x\}; S = \{x, x\}$

4. Consider a relation R(A,B,C,D,E) with dependencies ABC → DE, E → BCD. Then the relation R is in 3NF. State true or false with reasoning.

   Solution: False. ABC is a superkey of R. In case of the FD E → BCD, E is not a superkey and BCD is not part of the key. Hence the relation is not in 3NF.

Question 2. Relational algebra to English (20 points)

Consider a relation Speed (pitcher,fast,slow) recording the fastest (maximum) and slowest (minimum) pitch speeds for baseball pitchers. A key for the relation is (pitcher). Consider the following relational algebra expression, written in linear notation.

P1(pitcher,f) := $\pi_{pitcher,fast}$(Speed)
P2(pitcher,s) := $\pi_{pitcher,slow}$(Speed)
P3(pitcher) := $\pi_{pitcher}(P1 \bowtie_{f<fast}$ Speed)
P4(pitcher) := $\pi_{pitcher}(P2 \bowtie_{s>slow}$ Speed)
P5(pitcher) := $\pi_{pitcher}$(Speed) - P3
P6(pitcher) := $\pi_{pitcher}$(Speed) - P4
Answer(p) := P5 $\cup$ P6

State in English what is computed as the final Answer briefly. Long-winded answers will be deducted points.

Solution: Pitchers with the fastest and slowest pitch speeds.

Question 3. English to Relational algebra (20 points)

Consider a relation Taking(studentId,courseId) with no duplicates. Write a relational algebra expression to find all courses with at least five students. Your solution should be a single expression (i.e., not a tree or linear notation), and it should be as compact as possible while giving the correct answer on all databases.

Solution: $\pi_c\ \sigma_{(s1<s2\ AND\ s2<s3\ AND\ s3<s4\ AND\ s4<s5)}\ (\rho_{T1(s1,c)}(Taking) \bowtie \rho_{T2(s2,c)}(Taking) \bowtie \rho_{T3(s3,c)}(Taking) \bowtie \rho_{T4(s4,c)}(Taking) \bowtie \rho_{T5(s5,c)}(Taking))$

Question 4. Data to functional dependency (20 points)

Which of the following rules for functional dependencies are correct (i.e., the rule holds over all databases) and which are incorrect (i.e., the rule does not hold over some database)? For incorrect rules, give the simplest example relation instance you can come up with where the rule does not hold. For correct rules, prove them using Armstrong axioms.

1. If A $\rightarrow$ B and BC $\rightarrow$ D, then AC $\rightarrow$ D
2. If AB $\rightarrow$ C then A $\rightarrow$ C
3. If A $\rightarrow B_1, .., B_n$ and $C_1, .., C_m \rightarrow$ D and $C_1, ..., C_m$ is a subset of $B_1, .., B_n$, then A $\rightarrow$ D
4. If A $\rightarrow$ C and B $\rightarrow$ C and ABC $\rightarrow$ D, then A $\rightarrow$ D

Solution:

1. is correct
2. is incorrect - R(A,B,C) with R=(1,2,3),(1,4,5)
3. is correct
4. is incorrect - R(A,B,C,D) with R=(1,2,3,4),(1,5,3,6)

Question 5. Normalization (25 points)

Consider the following relational schema:

UnivInfo(studID, studName, course, profID, profOffice)

Each tuple in relation UnivInfo encodes the fact that the student with the given ID and name took the given course from the professor with the given ID and office. Assume that students have unique IDs but not necessarily unique names, and professors have unique IDs but not necessarily unique offices. Each student has one name; each professor has one office.

(a) Specify a set of completely nontrivial functional dependencies for relation UnivInfo that encodes the assumptions described above and no additional assumptions.

(b) Based on your functional dependencies in part (a), specify all minimal keys for relation UnivInfo.

(c) Is UnivInfo in Boyce-Codd Normal Form (BCNF) according to your answers to (a) and (b)? If not, give a decomposition of UnivInfo into BCNF.

(d) Now add the following two assumptions: (1) No student takes two different courses from the same professor; (2) No course is taught by more than one professor (but a professor may teach more than one course). Specify additional functional dependencies to take these new assumptions into account.

(e) Based on your functional dependencies for parts (a) and (d) together, specify all minimal keys for relation UnivInfo.

(f) Is UnivInfo in BCNF according to your answers to (d) and (e)? If not, give a decomposition of UnivInfo into BCNF.

Solution:

(a) studID $\rightarrow$ studName, profID $\rightarrow$ profOffice

(b) (studID,profID,course)

(c) No - neither studID nor profID is a key. Decomposition: R1(studID,studName), R2(profID,profOffice), R3(studID,course,profID)

(d) studID,profID $\rightarrow$ course, course $\rightarrow$ profID

(e) (studID,profID) (studID,course)

(f) No - studID, profID, and course are not keys. Decomposition: R1(studID,studName), R2(profID,profOffice), R3(studID,course), R4(course,profID)