

Proc SQL

Sue Erickson
Research Analyst
Office of Institutional Research
University of North Dakota



Office of Institutional Research

RRSUG 042408

Agenda

A basic review of using PROC SQL in SAS

- General SQL review
- Basic SQL query
- Multiple tables in SQL
- PROC SQL in SAS – general statement
- Using PROC SQL: DISTINCT, (*), COUNT, & SUM



Office of Institutional Research

RRSUG 042408

Structured Query Language

Simple query language using 3 basic steps

- SELECT *fields*
- FROM *tables*
- WHERE *criteria exist*
- ORDER BY *sort fields*



Office of Institutional Research

RRSUG 042408

Sample Tables

Employee_Table

Employee ID	Name	Hire Date	Job Title
4000	Jack	1/1/2004	Manager
4002	Jen	2/2/2005	Supervisor
4004	Jim	3/3/2006	Analyst
4006	Jody	4/4/2007	Assistant
4008	June	5/5/2008	Maintenance

Email_Table

Employee ID	Email Address
4000	Jack@eemail.com
4002	Jen@email.com
4004	Jim@email.com
4006	Jody@email.com
4008	June@email.com



Office of Institutional Research

RRSUG 042408

Simple Query

A query that will collect employees and their job descriptions if employee hired before 2007.

```
SELECT Employee_ID, Job_Title
FROM Employee_Table
WHERE Hire_Date < 01/01/2007
ORDER BY Job_Title
```

Employee ID	Name	Hire Date	Job Title
4000	Jack	1/1/2004	Manager
4002	Jen	2/2/2005	Supervisor
4004	Jim	3/3/2006	Analyst
4006	Jody	4/4/2007	Assistant
4008	June	5/5/2008	Maintenance



Office of Institutional Research

RRSUG 042408

Simple Query

```
SELECT Employee_ID, Job_Title
FROM Employee_Table
WHERE Hire_Date < 01/01/2007
ORDER BY Job_Title
```

Resulting output:

Employee ID	Job Title
4004	Analyst
4000	Manager
4002	Supervisor



Office of Institutional Research

RRSUG 042408

Multiple Table Query

```
SELECT A.Employee_ID, A.Name,
       B.Email_Address
FROM Employee_Table A,
     Email_Table B
WHERE A.Employee_ID=B.Employee_ID
```



Office of Institutional Research

RRSUG 042408

Multiple Table Query

```
SELECT A.Employee_ID, A.Name,
       B.Email_Address
FROM Employee_Table A, Email_Table B
WHERE A.Employee_ID=B.Employee_ID
```

Employee ID	Name	Email Address
4000	Jack	Jack@email.com
4002	Jen	Jen@email.com
4004	Jim	Jim@email.com
4006	Jody	Jody@email.com
4008	June	June@email.com



Office of Institutional Research

RRSUG 042408

ProcSQL in SAS

Will be using sample finance Data:

Unit	Account	DeptID	Fund	Project	Total Amt
UND01	516080	5110	31306		791.020
UND01	516005	2200	43500	UND0005155	0.000
UND01	621395	5300	20390		30.000
UND01	521075	2215	20568		1379.230
UND01	552095	5150	31300		336.100
UND01	693005	2052	31502		5367.350
UND01	516075	2400	40800	UND0010110	1.050
UND01	516170	1120	43500	UND0011284	153.240
UND01	542040	2315	20298		291.500
UND01	581060	2700	30205		75.960
UND01	490004	2365	30205		-22712.440
UND01	542040	8240	43300	UND0010118	858.740



Office of Institutional Research

RRSUG 042408

ProcSQL in SAS

```
PROC SQL;
CREATE table Expenses as
SELECT DeptID, Fund, Project, Account,
       Total_Amt
FROM FinanceData
WHERE Account >= '500000';
QUIT;
```



Office of Institutional Research

RRSUG 042408

ProcSQL in SAS

Expenses Table is created:

DeptID	Fund	Project	Account	Total_Amt
5110	31306		516080	791.02
2200	43500	UND0005155	516005	0
5300	20390		621395	30
2215	20568		521075	1379.23
5150	31300		552095	336.1
2052	31502		693005	5367.35
2400	40800	UND0010110	516075	1.05
1120	43500	UND0011284	516170	153.24
2315	20298		542040	291.5
2700	30205		581060	75.96
8240	43300	UND0010118	542040	858.74
2850	31203		611020	240



Office of Institutional Research

RRSUG 042408

How I use SAS Proc SQL - DISTINCT

```
PROC SQL;
CREATE table Funds_All as
SELECT Fund
FROM Expenses;
QUIT;
```

NOTE: Table
WORK.FUNDS_ALL
created, with 47825
rows and 1 column

```
PROC SQL;
CREATE table Funds_Distinct as
SELECT DISTINCT Fund
FROM Expenses;
QUIT;
```

NOTE: Table
WORK.FUNDS_DIST
INCT created, with
1486 rows and 1
column



Office of Institutional Research

RRSUG 042408

How I use SAS Proc SQL- DISTINCT

Generating a list of unique
Department/Fund Combinations:

```
PROC SQL;  
CREATE table Dept_Funds_Distinct as  
SELECT DISTINCT DeptID, Fund  
FROM Expenses;  
QUIT;
```



Office of Institutional Research

RRSUG 042408

How I use SAS Proc SQL - COUNT

"COUNT" is used to count the number of
occurrences for a specific criteria

```
PROC SQL;  
CREATE table FundCount as  
SELECT DeptID, COUNT(Fund) as Fund_Count  
FROM Expenses  
GROUP by DeptID;  
QUIT;
```

Table
WORK.FUNDCOUNT
created, with 210
rows and 2 columns.



Office of Institutional Research

RRSUG 042408

How I use SAS Proc SQL – WILD CARD (*)

"(*)" is used to include all lines in a table

```
PROC SQL;  
CREATE table DeptLineCount as  
SELECT DeptID, COUNT(*) as Line_Count  
FROM Expenses  
WHERE Total_Amt > 0  
GROUP by DeptID;  
QUIT;
```



Office of Institutional Research

RRSUG 042408

How I use SAS Proc SQL – WILD CARD (*)

In combination with DISTINCT, (*) can
allow user to select distinct lines from an
entire table.

```
PROC SQL;  
CREATE table DistinctLines as  
SELECT DISTINCT *  
FROM TableName;  
QUIT;
```



Office of Institutional Research

RRSUG 042408

How I use SAS Proc SQL - SUM

“SUM” is used to total a numeric field based on leading fields

PROC SQL;

```
CREATE table DeptFundExpenses as
SELECT DeptID, Fund, sum(Total_Amt) as Expense_Total
FROM FinanceData
WHERE Account >= '500000'
      AND Total_Amt ^ = 0
GROUP BY DeptID, Fund;
QUIT;
```



Office of Institutional Research

RRSUG 042408

SAS Proc SQL

Any questions???



Office of Institutional Research

RRSUG 042408