

Notes on Problem Set Solutions

Question 1.

For this question, we should count number of employees who response positively to the policy variables. We can use different measurement to catch employees' attitude toward them. First, we can count the number of people who say "very likely". We can also count the number of the who say "likely". We can even calculate percentage of people who response positively. We draw our conclusions based on the numbers and percentages. Second, notice that the response is ordered, with "0" stands for "very likely", with "1" stands for "likely" and "2" stands for not likely. We can actually average the number for each policy. The lower the averaged value, the more positive the response is. Most of you use the first approach. The query used for the second approach is presented below.

You may be familiar with query design view and hate the Sql script. To know what they look like in the query design view, you can copy them to your Access Sql window and then change to design view.

a. Convert data type of policy variable from text to number, and then run the following queries.
SELECT Count(PUB_DISC)/Count(surveyno) AS Expr1, Avg(EMPLOYEE.PUB_FLEX) AS AvgOfPUB_FLEX, Avg(EMPLOYEE.PUB_DAYCAR) AS AvgOfPUB_DAYCAR, Avg(EMPLOYEE.PUB_EMGNCY) AS AvgOfPUB_EMGNCY, Avg(EMPLOYEE.PUB_PKRATE) AS AvgOfPUB_PKRATE, Avg(EMPLOYEE.PUB_INFO) AS AvgOfPUB_INFO, Avg(EMPLOYEE.PUB_DISC) AS AvgOfPUB_DISC
FROM EMPLOYEE
WHERE (((EMPLOYEE.MODE1_TO) In ("1","2","3")));

With this query, we find policy associated with lowest value are discount passes and emergency rider services

b. for the four greatest four agencies, first find out them by the following query
SELECT Count(employee.SURVEYNO) AS CountOfSURVEYNO, employee.DOT_BRANCH
FROM employee
GROUP BY employee.DOT_BRANCH
ORDER BY Count(employee.SURVEYNO) DESC;
They are FAA, USCG, OST, FHWA
Then find out the most positively responded policies in these four agencies

```
SELECT Avg(EMPLOYEE.PUB_FLEX) AS AvgOfPUB_FLEX, Avg(EMPLOYEE.PUB_DAYCAR) AS AvgOfPUB_DAYCAR, Avg(EMPLOYEE.PUB_EMGNCY) AS AvgOfPUB_EMGNCY, Avg(EMPLOYEE.PUB_PKRATE) AS AvgOfPUB_PKRATE, Avg(EMPLOYEE.PUB_INFO) AS AvgOfPUB_INFO, Avg(EMPLOYEE.PUB_DISC) AS AvgOfPUB_DISC  
FROM EMPLOYEE  
WHERE (((EMPLOYEE.DOT_BRANCH) In ("FAA","USCG","OST","FHWA")) AND ((EMPLOYEE.MODE1_TO) In ("1","2","3")));
```

The most positively responded policies are the same as above.

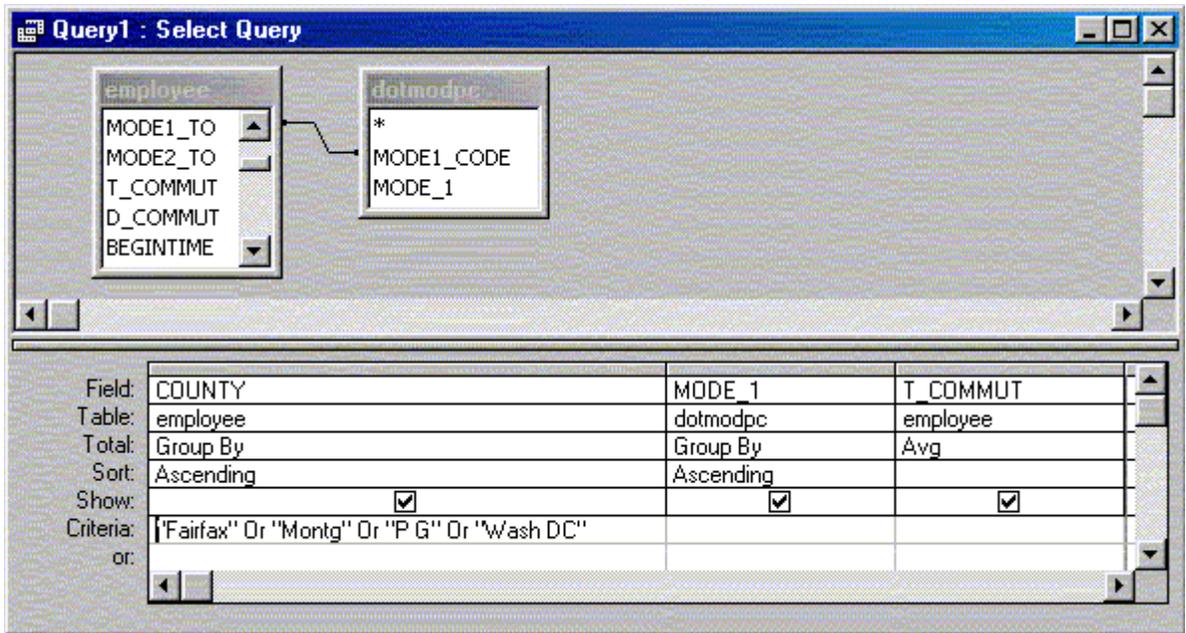
To make recommendation branch by branch, we can use the following query.

```
SELECT EMPLOYEE.DOT_BRANCH, Avg(EMPLOYEE.PUB_FLEX) AS AvgOfPUB_FLEX, Avg(EMPLOYEE.PUB_DAYCAR) AS AvgOfPUB_DAYCAR, Avg(EMPLOYEE.PUB_EMGNCY) AS AvgOfPUB_EMGNCY, Avg(EMPLOYEE.PUB_PKRATE) AS AvgOfPUB_PKRATE, Avg(EMPLOYEE.PUB_INFO) AS AvgOfPUB_INFO, Avg(EMPLOYEE.PUB_DISC) AS AvgOfPUB_DISC  
FROM EMPLOYEE  
WHERE (((EMPLOYEE.MODE1_TO) In ("1","2","3")))  
GROUP BY EMPLOYEE.DOT_BRANCH  
HAVING (((EMPLOYEE.DOT_BRANCH) In ("FAA","USCG","OST","FHWA")));
```

Almost all of you are OK with this questions. Some of you may understand the question a bit differently, it is OK with me when I graded it. You are given full credit or be fined only slightly. The main point is that you run the appropriate queries and get the important numbers. However, several students don't show the needed numbers to make the recommendation.

Question 2.

A. Table of alphabetical listing by COUNTY and MODE1 (name label) and Avg of T_COMMUTE:
The query design looks like this:



The SQL looks like this:

You may use commuting distance instead of commuting time. Both are OK. The table should look something like this:

COUNTY	MODE_1	AvgOfD_COMMUT
Fairfax	bicycle	43.6666666666667
Fairfax	carpool	40.1803797468354
Fairfax	commuterbus	59
Fairfax	commuterrail	47.7777777777778
Fairfax	drive alone	40.1030927835052
Fairfax	dropped off	45.3846153846154
Fairfax	metrobus	51.4905660377358
Fairfax	metrorail	55.0056818181818
Fairfax	motorcycle	28.75
Fairfax	other	50.9523809523809
Fairfax	vanpool	47.56
Montg	bicycle	43
Montg	carpool	44.9014084507042
Montg	commuterbus	90
Montg	commuterrail	67.12
Montg	drive alone	43.6666666666667
Montg	dropped off	41.25
Montg	metrobus	76.4285714285714
Montg	metrorail	55.6
Montg	motorcycle	30
Montg	other	91.6666666666667
Montg	vanpool	50.7142857142857
P G	bicycle	35
P G	carpool	40.0146750524109
P G	commuterbus	62.5
P G	commuterrail	57.2222222222222
P G	drive alone	34.8720930232558
P G	dropped off	38.59375
P G	metrobus	49.5844155844156
P G	metrorail	43.84375
P G	motorcycle	45
P G	other	38.75
P G	vanpool	49.7222222222222
Wash DC	bicycle	26.375
Wash DC	carpool	27.4444444444444
Wash DC	commuterbus	35
Wash DC	commuterrail	30
Wash DC	drive alone	20.3684210526316
Wash DC	dropped off	21.3793103448276
Wash DC	metrobus	36.6153846153846
Wash DC	metrorail	30.1532846715328
Wash DC	motorcycle	16.5
Wash DC	other	18
Wash DC	walk	15.1578947368421

B. The first and second most popular modes of travel in Montg are car pool and metro rail
SELECT dotmodpc.MODE_1, employee.COUNTY, Count(employee.SURVEYNO) AS
CountOfSURVEYNO, Avg(employee.D_COMMUT) AS AvgOfD_COMMUT
FROM dotmodpc INNER JOIN employee ON dotmodpc.MODE1_CODE = employee.MODE1_TO
GROUP BY dotmodpc.MODE_1, employee.COUNTY
HAVING (((employee.COUNTY)="Montg"))
ORDER BY employee.COUNTY, Count(employee.SURVEYNO) DESC;

For P G, the most popular ones are car pool and drive alone
SELECT dotmodpc.MODE_1, employee.COUNTY, Count(employee.SURVEYNO) AS
CountOfSURVEYNO, Avg(employee.D_COMMUT) AS AvgOfD_COMMUT
FROM dotmodpc INNER JOIN employee ON dotmodpc.MODE1_CODE = employee.MODE1_TO
GROUP BY dotmodpc.MODE_1, employee.COUNTY
HAVING (((employee.COUNTY)="P G"))
ORDER BY employee.COUNTY, Count(employee.SURVEYNO) DESC;

c. In Montg, 255 people have secondary commuting mode

```

SELECT Count(employee.SURVEYNO) AS CountOfSURVEYNO
FROM employee
WHERE (((employee.COUNTY)="Montg") AND ((employee.MODE2_TO)<>"0"));
For those without secondary commuting mode, the most popular primary commuting mode is
carpool ( 177 people carpool) .
SELECT dotmodpc.MODE_1, employee.COUNTY, Count(employee.SURVEYNO) AS
CountOfSURVEYNO, Avg(employee.D_COMMUT) AS AvgOfD_COMMUT, employee.MODE2_TO
FROM dotmodpc INNER JOIN employee ON dotmodpc.MODE1_CODE = employee.MODE1_TO
GROUP BY dotmodpc.MODE_1, employee.COUNTY, employee.MODE2_TO
HAVING (((employee.COUNTY)="Montg") AND ((employee.MODE2_TO)="0"))
ORDER BY employee.COUNTY, Count(employee.SURVEYNO) DESC;

```

Question 3.

A. County block groups with highest percentages of government workers:

Counties:	Tract:	Block group	Percentage of Government Workers:	Block group key
25013 Hampden	8114	3	100%	250138114003
25015 Hampshire	8203	4	63.6%	250158203004
25027 Worcester	7384	9	71.1%	250277384009

Note that full credit was given to those who had the answers in the table above. Government workers were determined by adding P0790003, P0790004, P0790005. The percentages were determined using the universe of employed persons as derived by adding up the fields P0790001, P0790002, P0790003, P0790004, P0790005, P0790006, P0790007. As seen in the Census description below, adding up the like categories provides an appropriate universe. There was confusion among some students who calculated percentages with the same universe used in Lab E. That is, the universe determined by adding male and females who worked in 1989. In the Census, work status has a different definition than employment status. While it was appropriate to use males and females who worked as the universe in the Lab E calculations, it produces wrong results when used with class of worker which is based on the universe of employed persons 16 years and over. See below.

P79. CLASS OF WORKER(7) [7]

Universe: Employed persons 16 years and over:

Private for profit wage and salary workers: P0790001

Private not for profit wage and salary workers: P0790002

Local government workers: P0790003

State government workers: P0790004

Federal government workers: P0790005

Self employed workers: P0790006

Unpaid family workers: P0790007

Because the confusion was due to these subtle but important differences, we gave credit to answers based on the universe of workers that was used in Lab E. However, full credit was given to the correct answers.

Question 4.

Many students did well on this question and correctly chose Map 2 as the most important map to use in the analysis. Note that Map 2 is the most appropriate map because it shows the relative density of adults without degrees. Normalizing the data (e.g., determining persons per square kilometer) provides a standard way to compare regions. However, the argument on highest density alone is not enough. We should also think of the density of the neighbors. Think of whether it is easily accessible from other towns. That's a very important reason we use maps instead of using only the underlying tables.