**UJWALA POLEPEDDI**

**TEACH ONE- STRATIFIED COVARIATE BALANCING**

**QUESTION 2:**

The following data provide the survival among stomach cancer patients.  The data provides 35 common comorbidities for patients who have or don't have stomach cancer.

1. Using SQL, group the diagnoses into commonly occurring strata.
2. Within each strata, calculate the odds of mortality from cancer.
3. Calculate the common odds ratio across strata.
4. Conduct sensitivity analysis for the calculated common odds ratio.  Sensitivity analysis is the process of changing one variable and re-examining the conclusions. Drop one of the 35 comorbidities from the analysis and repeat the entire analysis and check that 65% of cases are matched to controls. The percent of cases that are matched is called overlap.  It is defined as:  
     
     
     
   In most problems, one wants to maximize the overlap to be around at least 80%, so that findings can be generalized to the original cases.

Report how the un-confounded and confounded odds of mortality from stomach cancer are different from each other?   [**Data►**](http://openonlinecourses.com/causalanalysis/StomachCancer.csv)[**Using Synthetic Controls►**](http://openonlinecourses.com/causalanalysis/Synthetic%20Controls%20Code.docx) [**Using Parents in Markov Blanket►**](http://openonlinecourses.com/causalanalysis/Markov%20Balnket%20for%20Stratified%20Covariates.docx)

**First group the diagnosis into commonly occurring strata**

SELECT ([Column 0])as ID

,cast([Cancer] as int)as CA

,[I305 1]

,[I309 81]

,[I311 ]

,[IE849 7]

,[I150 9]

,[I276 1]

,[I276 8]

,[I530 81]

,[I263 9]

,[I276 51]

,[IV15 82]

,[I511 9]

,[I401 9]

,[I787 20]

,[I564 00]

,[I272 4]

,[I280 9]

,[I285 9]

,[I496 ]

,[I458 9]

,[I486 ]

,[IV58 61]

,[I197 7]

,[I578 9]

,[I584 9]

,[IV66 7]

,[I244 9]

,[I414 01]

,[I599 0]

,[I414 00]

,[I585 9]

,[I600 00]

,[I428 0]

,[I427 31]

,[I403 90]

,cast([Dead] as int) as Dead

into stoamch\_cancer1 FROM [Stratified\_covariate].[dbo].[StomachCancer]-----(100000 rows affected)

Select ID,CA,[I305 1]+[I309 81]+[I311 ]+[IE849 7]+[I150 9]+[I276 1] +[I276 8] +[I530 81]+[I263 9]

+[I276 51]+[IV15 82]+[I511 9]+[I401 9]+[I787 20]+[I564 00]+[I272 4]+[I280 9]+[I285 9]

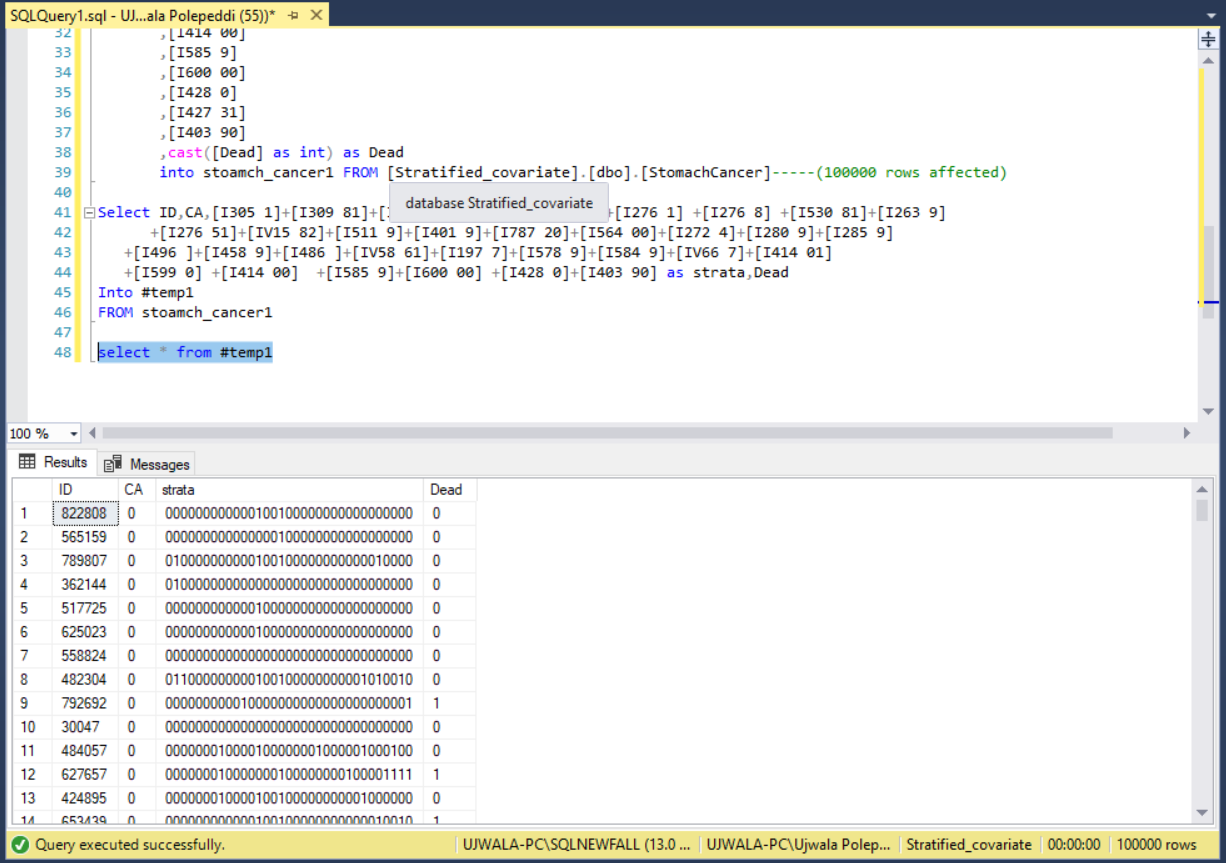
+[I496 ]+[I458 9]+[I486 ]+[IV58 61]+[I197 7]+[I578 9]+[I584 9]+[IV66 7]+[I414 01]

+[I599 0] +[I414 00] +[I585 9]+[I600 00] +[I428 0]+[I403 90] as strata,Dead

Into #temp1

FROM stoamch\_cancer1

select \* from #temp1



--CASES

---drop table #cases

SELECT COUNT(distinct [ID]) AS nCases -- Number of residents unable to eat

, Sum(IIF([Dead] = 1, 1., 0.)) AS a

, SUM(IIF([Dead] = 0, 1., 0.)) AS b

, strata

INTO #Cases -- Save in temporary file called Cases

FROM #temp1

WHERE [CA] = 1 -- Select only residents who were unable to eat

GROUP BY strata -------(176 rows affected)

--CONTROLS

--drop table #controls

SELECT COUNT(distinct [ID]) AS nControls -- Number of residents unable to eat

, Sum(IIF([Dead] = 1, 1., 0.)) AS c

, SUM(IIF([Dead] = 0, 1., 0.)) AS d

, strata

INTO #Controls -- Save in temporary file called Cases

FROM #temp1

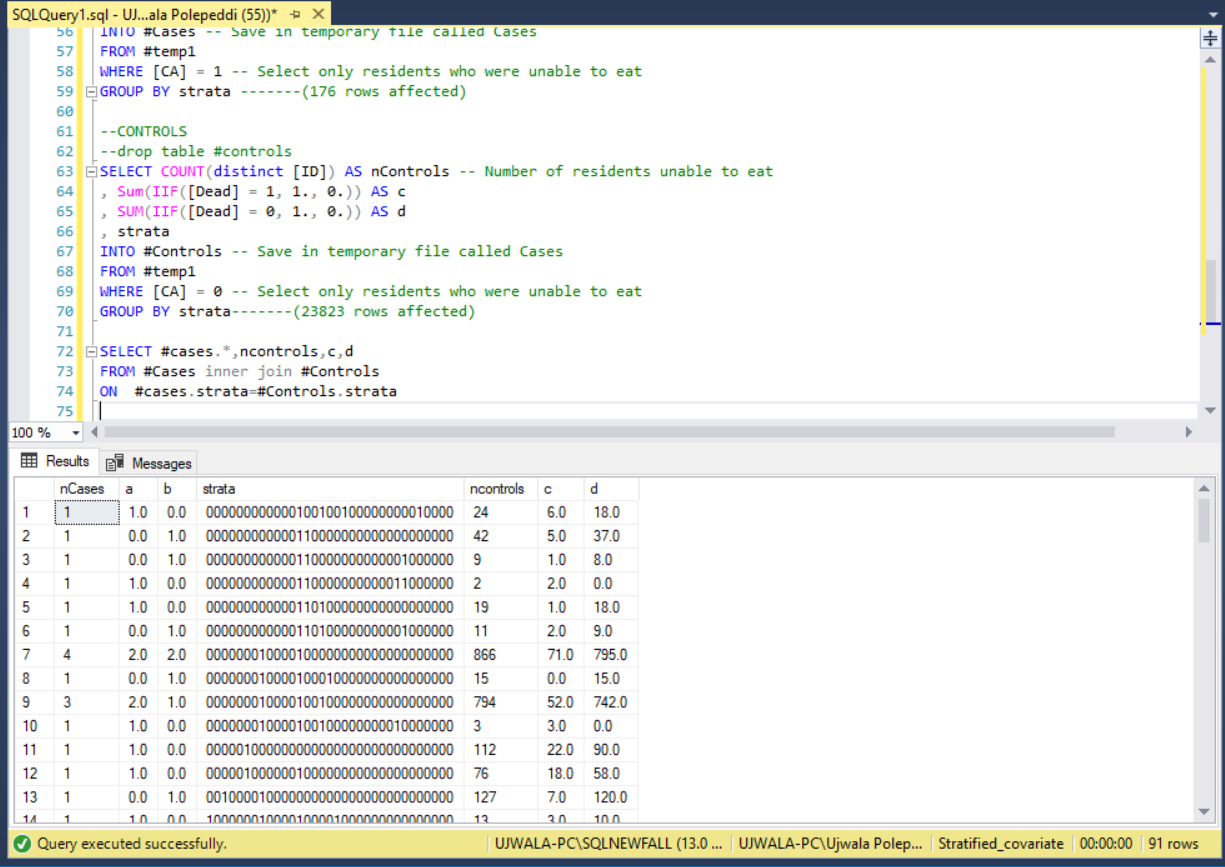
WHERE [CA] = 0 -- Select only residents who were unable to eat

GROUP BY strata-------(23823 rows affected)

SELECT #cases.\*,ncontrols,c,d

FROM #Cases inner join #Controls

ON #cases.strata=#Controls.strata



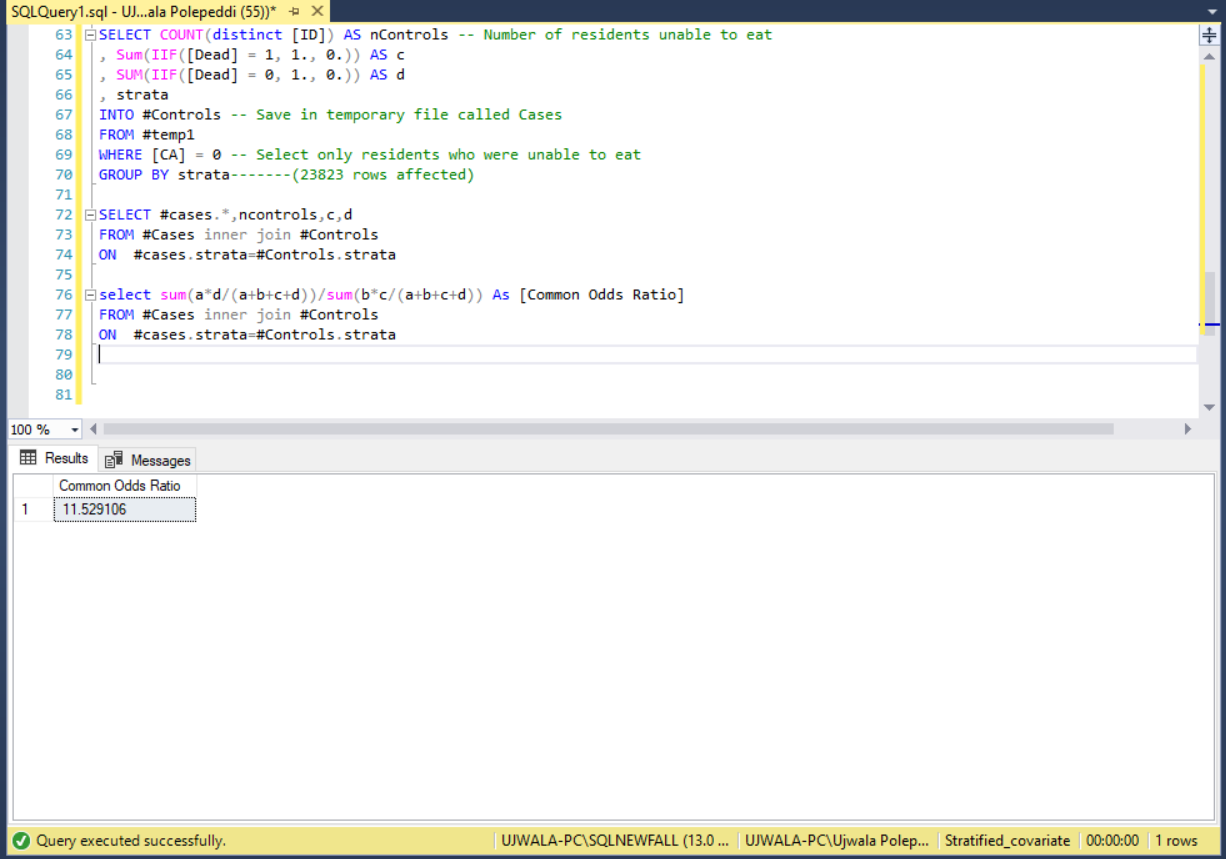
**Within each strata, calculate the odds of mortality from cancer**

-----CALCULATING ODDS RATIO

select sum(a\*d/(a+b+c+d))/sum(b\*c/(a+b+c+d)) As [Common Odds Ratio]

FROM #Cases inner join #Controls

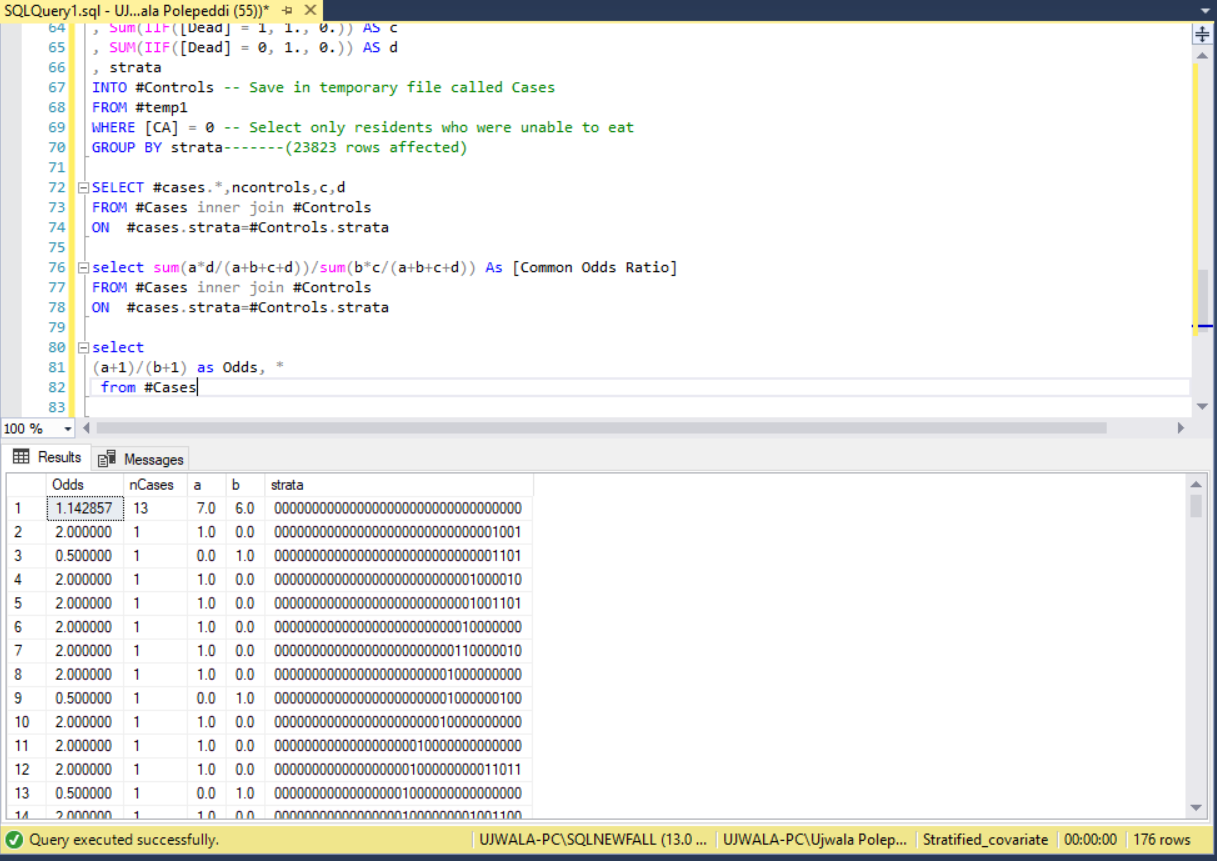
ON #cases.strata=#Controls.strata



select

(a+1)/(b+1) as Odds, \*

from #Cases



**Calculate the common odds ratio across strata.**

Declare @TotalCases Float

Set @TotalCases = (Select Count ([Dead]) from #temp1 Where [CA] =1)

SELECT sum(cast((a\*d)/(a+b+c+d)as float))/sum(cast((b\*c)/(a+b+c+d)as float)) As [Common Odds]

,sum (cast((a+b)/(c+d)as float)) As WeightControl

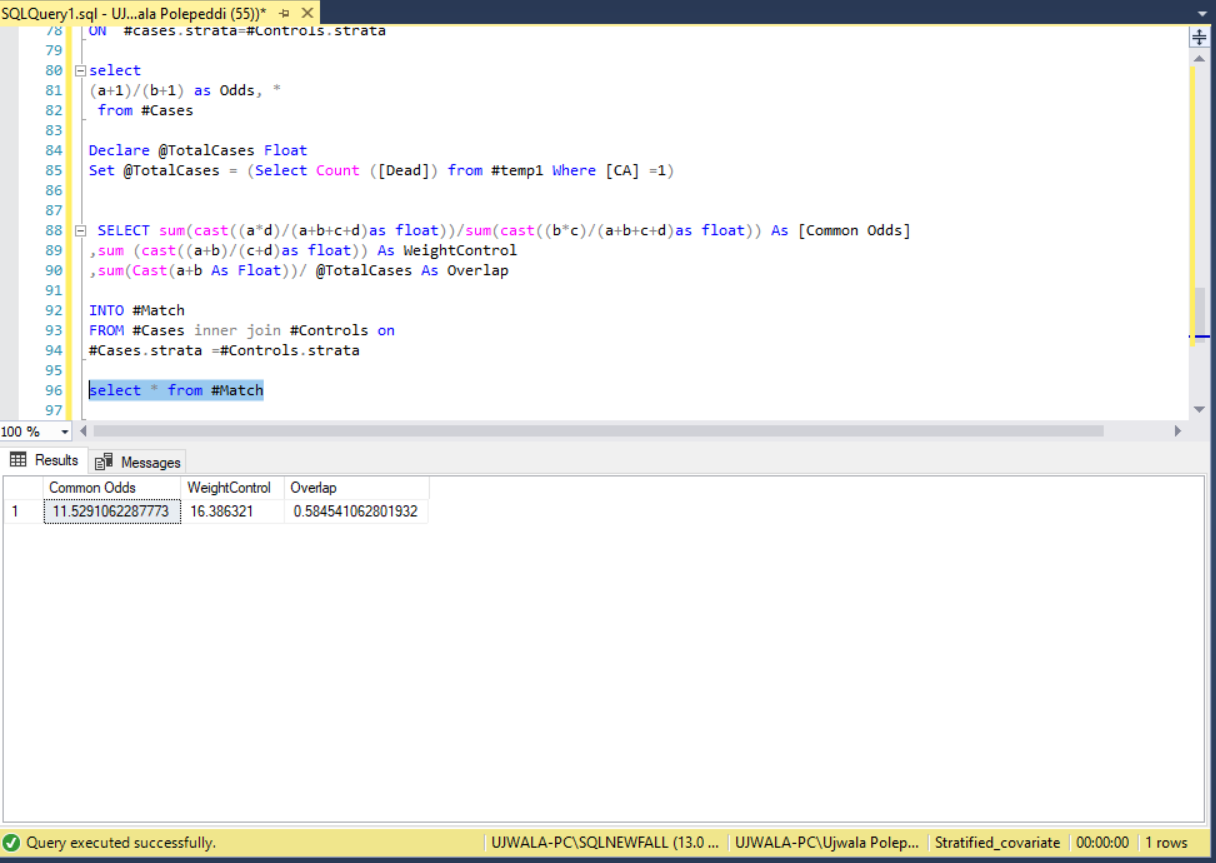
,sum(Cast(a+b As Float))/ @TotalCases As Overlap

INTO #Match

FROM #Cases inner join #Controls on

#Cases.strata =#Controls.strata

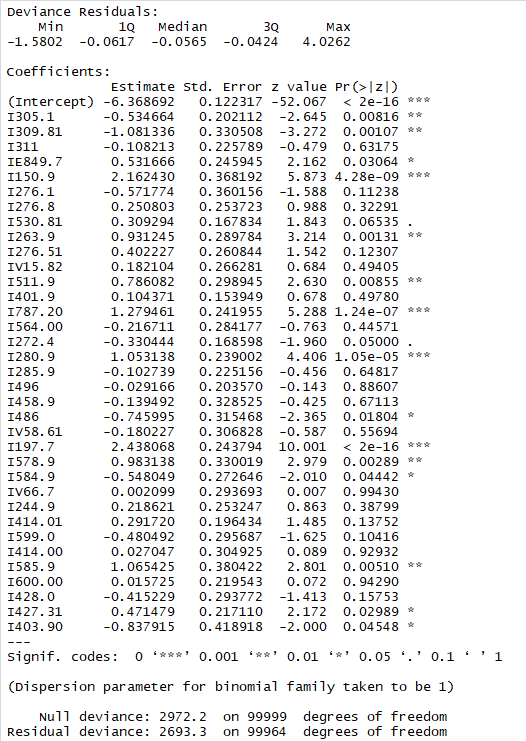
select \* from #Match



**Conduct sensitivity analysis for the calculated common odds ratio.**

**Sensitivity analysis is the process of changing one variable and re-examining the conclusions.**

**LOADING AND RUNNING THE DATA IN R TO FIND SIGNIFICANT VARIABLES**

****

Taking into consideration the significant variables and dropping the rest.

Re-run the SQL code for calculating the common odds ratio and percentage overlap.

Drop table #Cases

SELECT COUNT(distinct ID) AS nCases

, Sum(IIF([Dead] = 1, 1., 0.)) AS a

, SUM(IIF([Dead] = 0, 1., 0.)) AS b

,[I305 1], [I309 81]

, [IE849 7], [I150 9], [I530 81], [I263 9]

, [I787 20], [I272 4], [I280 9], [I486 ]

, [I578 9], [I584 9]

, [I585 9], [I427 31], [I403 90]

INTO #Cases

FROM stoamch\_cancer1

WHERE [CA] = 1

GROUP BY [I305 1], [I309 81]

, [IE849 7], [I150 9], [I530 81], [I263 9]

, [I787 20], [I272 4], [I280 9], [I486 ]

, [I578 9], [I584 9]

, [I585 9], [I427 31], [I403 90]-------(95 rows affected)

--- Grouping Controls that have NO Cancer by the conditions & seperating dead to column A & B

Drop table #Controls

SELECT COUNT(distinct [ID]) AS nControls

, Sum(IIF([Dead] = 1, 1., 0.)) AS c

, SUM(IIF([Dead] = 0, 1., 0.)) AS d

, [I305 1], [I309 81]

, [IE849 7], [I150 9], [I530 81], [I263 9]

, [I787 20], [I272 4], [I280 9], [I486 ]

, [I578 9], [I584 9]

, [I585 9], [I427 31], [I403 90]

INTO #Controls

FROM stoamch\_cancer1

WHERE [CA] = 0

GROUP BY [I305 1], [I309 81]

, [IE849 7], [I150 9], [I530 81], [I263 9]

, [I787 20], [I272 4], [I280 9], [I486 ]

, [I578 9], [I584 9]

, [I585 9], [I427 31], [I403 90] --(2493 rows affected)

-- Calculating the Common Odd Ratio

SELECT sum(a\*d/(a+b+c+d))/sum(b\*c/(a+b+c+d)) As [Common Odds Ratio]

into #OddsRatio

FROM #Cases inner join #Controls

ON #Cases.[I305 1] =#Controls.[I305 1]

and #Cases.[I309 81] = #Controls.[I309 81]

and #Cases.[IE849 7]= #Controls.[IE849 7]

and #Cases.[I150 9]= #Controls.[I150 9]

and #Cases.[I530 81]= #Controls.[I530 81]

and #Cases.[I263 9]= #Controls.[I263 9]

and #Cases.[I787 20]= #Controls.[I787 20]

and #Cases.[I272 4]= #Controls.[I272 4]

and #Cases.[I280 9]= #Controls.[I280 9]

and #Cases.[I486 ]= #Controls.[I486 ]

and #Cases.[I578 9]= #Controls.[I578 9]

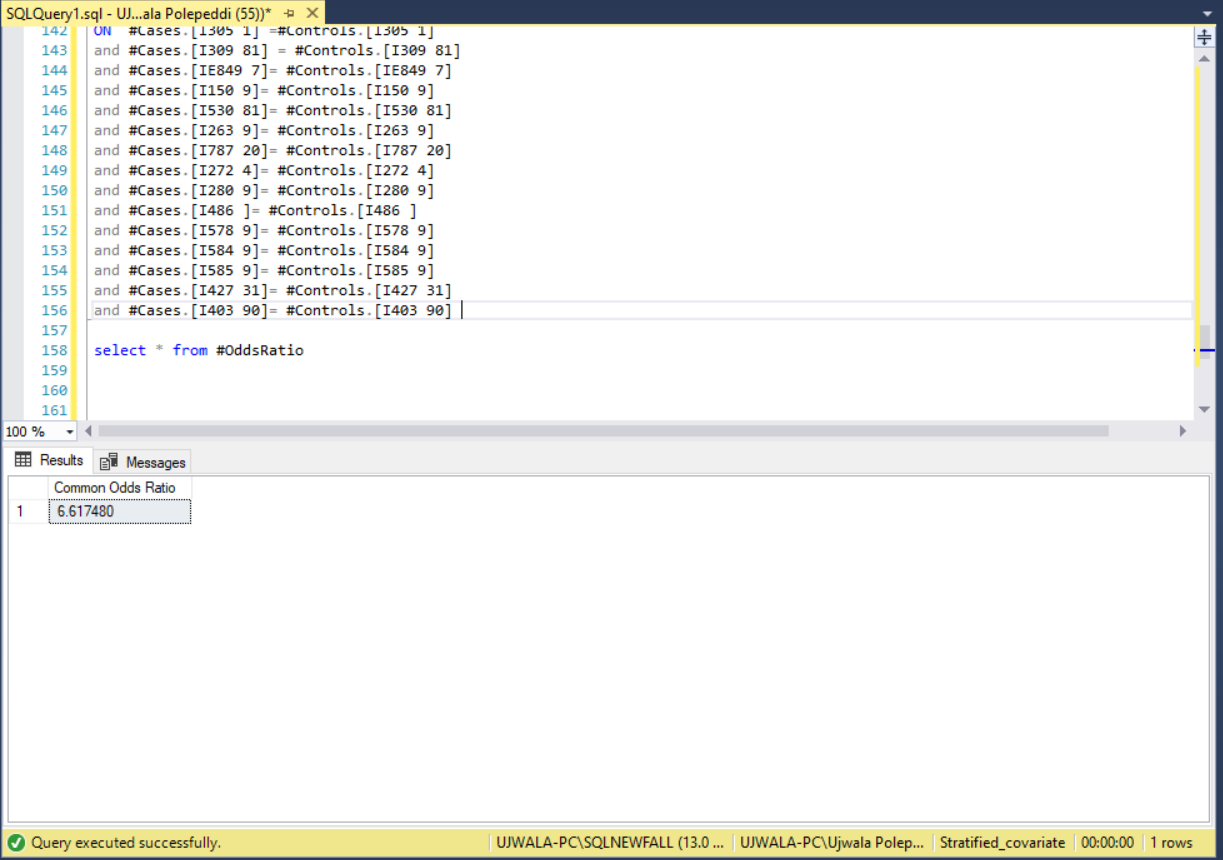
and #Cases.[I584 9]= #Controls.[I584 9]

and #Cases.[I585 9]= #Controls.[I585 9]

and #Cases.[I427 31]= #Controls.[I427 31]

and #Cases.[I403 90]= #Controls.[I403 90]

select \* from #OddsRatio



-- Matching Cases with controls

drop table #Match

select a ,b ,c ,d into #Match

from #Cases join #Controls

ON #Cases.[I305 1] =#Controls.[I305 1]

and #Cases.[I309 81] = #Controls.[I309 81]

and #Cases.[IE849 7]= #Controls.[IE849 7]

and #Cases.[I150 9]= #Controls.[I150 9]

and #Cases.[I530 81]= #Controls.[I530 81]

and #Cases.[I263 9]= #Controls.[I263 9]

and #Cases.[I787 20]= #Controls.[I787 20]

and #Cases.[I272 4]= #Controls.[I272 4]

and #Cases.[I280 9]= #Controls.[I280 9]

and #Cases.[I486 ]= #Controls.[I486 ]

and #Cases.[I578 9]= #Controls.[I578 9]

and #Cases.[I584 9]= #Controls.[I584 9]

and #Cases.[I585 9]= #Controls.[I585 9]

and #Cases.[I427 31]= #Controls.[I427 31]

and #Cases.[I403 90]= #Controls.[I403 90] --(88 rows affected)

The sum of the controls and cases 73804.0

select \* from #Match

--Calculating the overalp

Declare @TotalCases Float

SET @totalCases = (SELECT Sum(nCases) FROM #Cases)

SELECT ROUND(SUM(a+b)\*100/@TotalCases,2) as [Percent Overlap]

FROM #Match

