**Survival after Disabilities**

**Question 3:**  In a nursing home, data were collected on residents' survival and disabilities.  The data are listed in the following order: ID, age, gender (M for male, F for Female), number of assessments completed on the person, number of days followed, days since first assessment, days to last assessment, unable to eat, unable to transfer, unable to groom, unable to toilet, unable to bathe, unable to walk, unable to dress, unable to bowel, unable to urine, dead (1) or alive (0), and assessment number.  Analyze these data using the following steps” Provide both SQL code and results in the space provided. Here are the data for this question:  [Data►](http://openonlinecourses.com/causalanalysis/Assessments.zip)

1. Calculate age of assessment as age + [Days from First Assessment]/365. Calculate for each assessment whether within 6 months of the assessment the individual dies. Eliminate from analysis all assessments that occur after the individual was unable to toilet. (5 points)

Ans:-

SELECT [Column 0] as ID,

 Cast([Column 1] AS Float) + CAST([Column 5] AS Float)/365. AS [Age]

 ,IIF ([Column 2]='M', 1, 0) AS [Male]

 ,[Column 3] AS [assessments\_completed]

 ,[Column 4] AS [Days followed],CAST([Column 5] AS Float) AS [Days\_first\_assessment]

 ,[Column 6] AS [days\_last\_assessment]

 ,[Column 7] AS [UEat]

 ,[Column 8] AS [UTransfer]

 ,[Column 9] AS [UGroom]

 ,[Column 10] AS [UToilet]

 ,[Column 11] AS [UBathe]

 ,[Column 12] AS [UWalk]

 ,[Column 13] AS [UDress]

 ,[Column 14] AS [UBowel]

 ,[Column 15] AS [UUrine]

 ,[Column 16] AS [Dead]

 ,[Column 17] AS [Assessment\_no]

 INTO #Dataq3

 FROM [HAP 823 EXAM].[dbo].[Assessments] --(1306456 rows affected)

 -- DEAD PATIENTS IN 6 MONTHS

 SELECT ID

, [Days\_first\_assessment] AS [Dead\_day]

INTO #Dead

 FROM #Dataq3 WHERE Dead=1 --(196533 rows affected)

SELECT #Dataq3.\*

, IIF(#dead.[Dead\_day]-#Dataq3.[Days\_first\_assessment]

between 0 and 183, 1,0) AS [Dead6m]

INTO #Dead6m FROM #Dataq3 left join #Dead ON #Dataq3.[ID]=#Dead.[ID] --(1306456 rows affected)

--for unable to toilet

select ID ,

min ([Days\_first\_assessment]) AS UToilet\_Day

into #UT from #Dataq3

where UToilet=1 group by ID --(124375 rows affected)

Select MIN(Age) AS Min\_Age INTO #MIN FROM #Dead6m GROUP BY ID

Declare @AvgAge AS Float

SET @AvgAge=(SELECT Avg(Min\_Age) FROM #MIN)

SELECT #Dead6m.ID, [Assessment\_no], Dead6M, [UToilet]

, (IIF(Age<@AvgAge,'1', '0'))as avg\_age,Male

, UTransfer,UGroom,UEat,UBathe,UWalk,UDress,UBowel,UUrine

into dataq3R FROM #Dead6m left join #UT ON #Dead6m.[ID]=#UT.[ID]

WHERE #Dead6m.[Days\_first\_assessment]<=#UT.UToilet\_Day or #UT.UToilet\_Day is null

/\*(254732 rows affected) (769638 rows affected) \*/

 select \* from dataq3R

export this data into R and regress for Question b

1. Identify parents in Markov Blanket (pMB) of unable to toilet. Regress unable to toilet on prior events. Variables that are significant at levels less than 0.05 and have an effect size larger than 0.4 or smaller than -0.4 are on the pMB. (10 points)

Ans:-

> dataq3r=read.csv('dataq3r.csv')

> colnames(dataq3r)

 [1] "ID" "Assessment\_no" "Dead6M" "UToilet" "avg\_age"

 [6] "Male" "UTransfer" "UGroom" "UEat" "UBathe"

[11] "UWalk" "UDress" "UBowel" "UUrine"

> modelq3r=glm(dataq3r[,4]~.,data=dataq3r[,3] + dataq3r[,(5:14)] , family=binomial)

> summary(modelq3r)

Call:

glm(formula = dataq3r[, 4] ~ ., family = binomial, data = dataq3r[,

 3] + dataq3r[, (5:14)])

Deviance Residuals:

 Min 1Q Median 3Q Max

-3.7052 -0.2721 -0.1604 -0.0875 3.4674

Coefficients:

 Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.865220 0.015025 -57.585 < 2e-16 \*\*\*

avg\_age -0.968282 0.009504 -101.887 < 2e-16 \*\*\*

Male -3.729115 0.015718 -237.249 < 2e-16 \*\*\*

UTransfer 0.047720 0.016703 2.857 0.00428 \*\*

UGroom 1.143939 0.015547 73.578 < 2e-16 \*\*\*

UEat -0.446403 0.016346 -27.310 < 2e-16 \*\*\*

UBathe 0.190747 0.018010 10.591 < 2e-16 \*\*\*

UWalk 1.024939 0.011051 92.743 < 2e-16 \*\*\*

UDress 3.017073 0.011466 263.138 < 2e-16 \*\*\*

UBowel 1.134752 0.014947 75.919 < 2e-16 \*\*\*

UUrine -0.099163 0.013942 -7.113 1.14e-12 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

 Null deviance: 680848 on 769637 degrees of freedom

Residual deviance: 308837 on 769627 degrees of freedom

AIC: 308859

Number of Fisher Scoring iterations: 7

To stratify on MARKOV FOR all significance <0.05 and effect size larger than 0.4 or smaller than -0.4 are:

 Avg\_age, Male, UGroom, UEat, Uwalk, UDress, UBowel, UUrine

1. Stratify the variables on pMB and calculate the impact of unable to toilet on survival. (10 points)

Ans:-

SELECT COUNT([ID]) AS nCases

,Sum(IIF([Dead6M] = 1, 1., 0.)) AS DeadCase

, SUM(IIF([Dead6M] = 0, 1., 0.)) AS AliveCase

, UToilet,avg\_age,MALE,UGroom,UEat,UWalk,UDress,UBowel,UUrine

INTO #Cases FROM dataq3R

WHERE UToilet = 1-- Select only assessments unable to toilet

GROUP BY avg\_age,MALE,UGroom,UEat,UWalk,UDress,UBowel,UUrine, UToilet --(203 rows affected)

 -- controls for able to toilet

 SELECT COUNT([ID]) AS nControls

,Sum(IIF([Dead6M] = 1, 1., 0.)) AS DeadControls

, SUM(IIF([Dead6M] = 0, 1., 0.)) AS AliveControls

, UToilet,avg\_age,MALE,UGroom,UEat,UWalk,UDress,UBowel,UUrine

INTO #Controls FROM dataq3R

WHERE UToilet = 0-- Select only assessments able to toilet

GROUP BY avg\_age,MALE,UGroom,UEat,UWalk,UDress,UBowel,UUrine, UToilet --(200 rows affected)

-- to match case with controls

select #Cases.\*,ncontrols,DeadControls,AliveControls

into #M3

FROM #Cases INNER JOIN #Controls ON

#Cases. avg\_age = #Controls. avg\_age AND

#Cases. MALE = #Controls. MALE AND

#Cases. UGroom = #Controls. UGroom AND

#Cases. UEat = #Controls. UEat AND

#Cases. UWalk = #Controls. UWalk AND

#Cases. UDress = #Controls. UDress AND

#Cases. UBowel = #Controls. UBowel AND

#Cases. UUrine = #Controls. UUrine --- (186 rows affected)

1. Report the overlap between cases, unable to toilet, and controls, able to toilet (5 points).

Ans:-

-- for odd ratio

SELECT sum(DeadCase\*AliveControls/(DeadCase+AliveCase+DeadControls+AliveControls))

/sum(AliveCase\*DeadControls/(DeadCase+AliveCase+DeadControls+AliveControls)) As CommonOddRatio

FROM #M3 /\*CommonOddRatio 1.397327 \*/

-- for overlap

Declare @TotalCases Float

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

SELECT ROUND(SUM(nCases)\*100/@TotalCases,2) as overlapPercent
FROM #M3 /\*overlapPercent 99.69 \*/