\* reading in the data from csv format

import delimited C:\Users\dhnsingh\Documents\HealthScoreExercise\Acumen\_Data\_Analysis\_Exercise.csv

\* renaming header for first col

rename ïobservationnumber observationnumber

label variable observationnumber "Observation Number"

browse

describe // 19,000 observations

notes

\*\*\*

\* Q1: Understanding the Data

\* variable ranges

sum

\* missing values

search mdesc

mdesc

\* demographics over time

\* (i) race distribution per quarter

graph bar (count), ///

over(race,label(labsize(small))) ///

over(quarter, label(labsize(small))) ///

ytitle("Count by Race", size(small)) ///

title("Race Distribution over Time") asyvars

\* (ii) sex distribution per quarter

graph bar (count), ///

over(sexmale1,label(labsize(small))) ///

over(quarter, label(labsize(small))) ///

ytitle("Count by Sex", size(small)) ///

title("Sex Distribution over Time") asyvars

\* (iii) sex distribution per quarter

graph box age if 20<age & age<60, ///

over(quarter, label(labsize(small))) ///

ytitle("Age", size(small)) ///

title("Age Range over Time")

graph bar (count) if 20<age & age<60, ///

over(age,label(labsize(small))) ///

by(quarter) ///

ytitle("Count by Age", size(small)) ///

title("Age Distribution over Time")

\* Q2: Evaluating the Claim

sum healthscore, d

correlate

pwcorr quarter sexmale1 race age hospitalvisitthisquarter1yes salary healthscore, print(0.05) star(0.01)

\* binning data

egen healthcat = cut(healthscore), at(0, 1, 2, 3, 4, 5, 6, 11)

tab healthcat

tab healthcat quarter, col

\* health category distribution over time

graph bar (count), ///

over(healthcat,label(labsize(small))) ///

over(quarter, label(labsize(small))) ///

ytitle("Count by Healthscore", size(small)) ///

title("Healthscore Distribution over Time") asyvars

\* health category by hospital vissts

graph bar (count), ///

over(hospitalvisitthisquarter1yes,label(labsize(small))) ///

over(healthcat, label(labsize(small))) ///

ytitle("Count by Healthscore", size(small)) ///

title("Healthscore Distribution by Hospital Visits") asyvars

\* plotting continuous variables using binned data

search binscatter

binscatter healthscore age

binscatter healthscore salary

graph7 healthscore, bin(6) normal f

\* Q3: Exploring Relationships

xtset employeeid quarter

bysort quarter: egen healthscore\_mean1=mean(healthscore)

twoway scatter healthscore quarter if healthscore<7, msymbol(circle\_hollow) || connected healthscore\_mean1 quarter,

msymbol(diamond) || , xlabel(1(1)12)

\* model 1: time and entity fe

xtreg healthscore i.quarter sexmale1 race age hospitalvisitthisquarter1yes salary, fe

\* model 2: time and entity fe

xi: areg healthscore i.quarter sexmale1 race age hospitalvisitthisquarter1yes salary, absorb(employeeid)

\* model 3: random effects

xtreg healthscore sexmale1 race age hospitalvisitthisquarter1yes salary, re robust