Detailed Report

## Code for initializing data and looking into the basic statistical insights

import numpy as np

import csv

import pandas as pd

import matplotlib.pyplot as plt

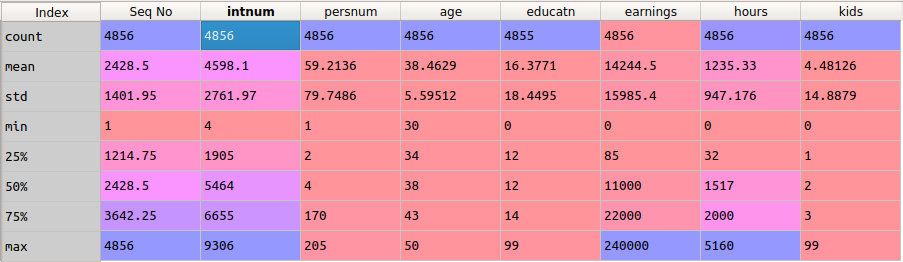
#%matplotlib inline

datafile = "PSID.csv"

D1=pd.read\_csv(datafile,delimiter=',')

Atrr=D1.describe()

## Sample Statistics



## Analyzing the age distribution

import csv

import pandas as pd

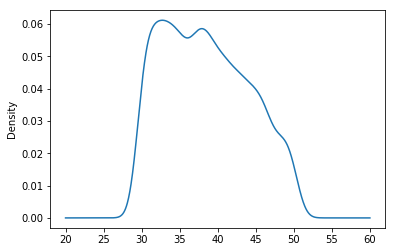
import matplotlib.pyplot as plt

#%matplotlib inline

datafile = "PSID.csv"

D1=pd.read\_csv(datafile,delimiter=',')

D1.age.plot(kind='kde')



The age distribution of the sample is biased towards the 30-40 range.

* From the sample statistics age has shown biased to 38.46 with the standard deviation of 5.59.
* The above distribution plot for age shows that the age is not uniformly distributed.

## Hypothesis/Question

Confidence interval is given as  
CI=statistic +- z\*SE

Where as  
z=(statistic-H0)/sigma  
SE=sigma/root(n)

1. **"Is education level affect to keep the marriage?"**

* People who have education level higher than level 10 have high probability of ended with divorce/seperated compared to the rest.

µs – Mean of education level for separated or divorced individuals

µn- Mean of education level for married and non-separated individuals

H0 => Null Hypothesis µs= µn  
Ha => Alternate Hypothesis µs ≠ µn

MarriedOnly=D1.loc[D1['married'] == 'married']

µn=MarriedOnly['educatn'].mean() #mean education level for married

Seperated=D1.loc[D1['married']=='separated']# mean education level for seperated/divorced

Seperated= D1.loc[D1['married']=='divorced']

µs=Seperated['educatn'].mean()

#confidence interval calculation

statistic1=D1['educatn'].mean()

sigma1=D1['educatn'].std()

z1=(statistic1-0)/(sigma1)

SE1=sigma1/math.sqrt(D1['educatn'].count())

CI1=(statistic1+z1\*SE),(statistic1-z1\*SE1)

µs=15.33  
µn=19.34 Since µs<µn alternate hypothesis holds and data is statistically significant.  
Hence increase in education level increases seperation/ divorce :).

Confidence Interval is (16.612,16.142)

1. **"Is working more hours increase the average income?"**

µh – Mean of average income for individuals working more than 2600 hours  
µl- Mean of average income for individuals working less than 2600 hours

H0 => Null Hypothesis µh= µl  
Ha => Alternate Hypothesis µh> µl

Highworkinghours=D1.loc[D1['hours'] >= 2600]

µh=Highworkinghours['earnings'].mean()

Lowworkinghours=D1.loc[D1['hours']<2600]# or

µl=Lowworkinghours['earnings'].mean()

#confidence interval calculation

statistic2=D1['earnings'].mean()

sigma2=D1['earnings'].std()

z2=(statistic2-0)/(sigma2)

SE2=sigma2/math.sqrt(D1['earnings'].count())

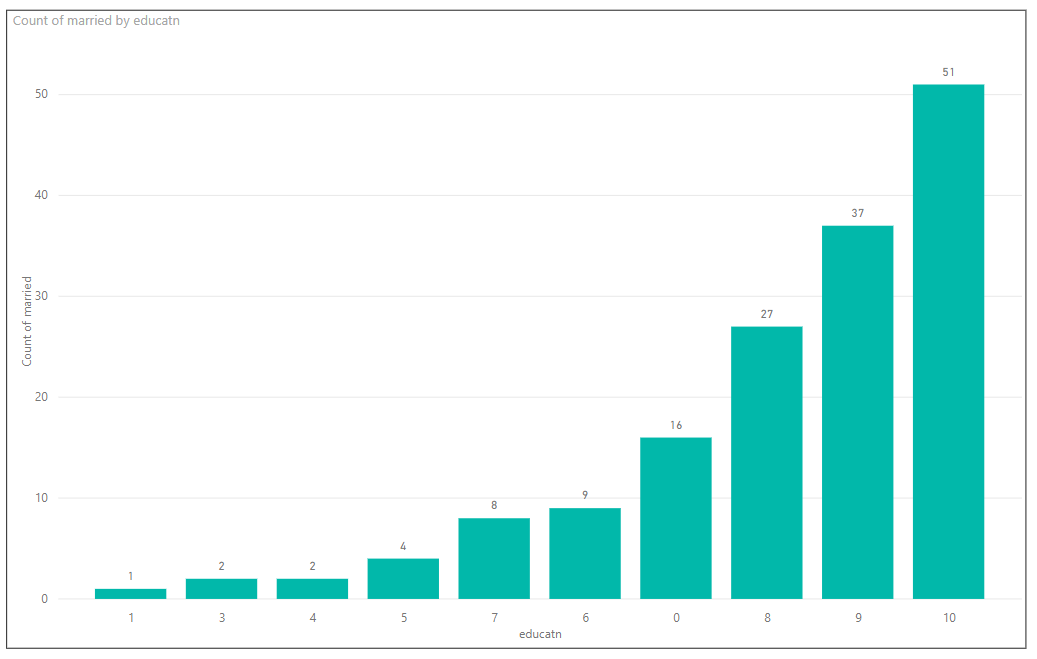
CI2=(statistic2+z2\*SE),(statistic2-z2\*SE2)

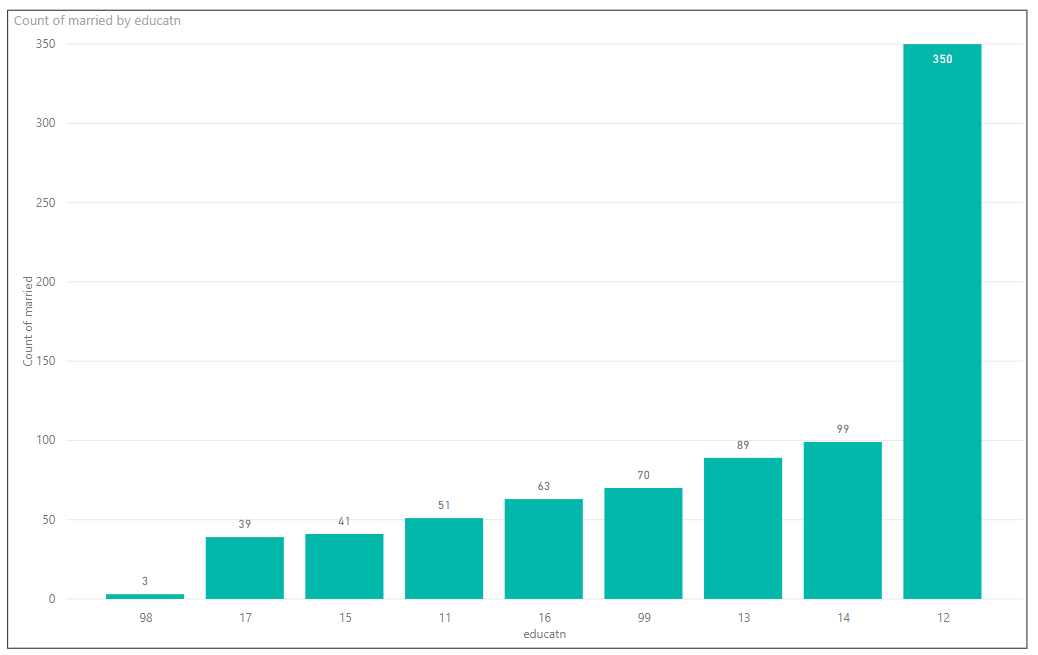
µh=30955.48 µl=13564.03

Since µl<µh alternate hypothesis holds and data is statistically significant.  
Hence increase in working hours increases earnings.  
Confidence Interval is (14244.74,14040.09)

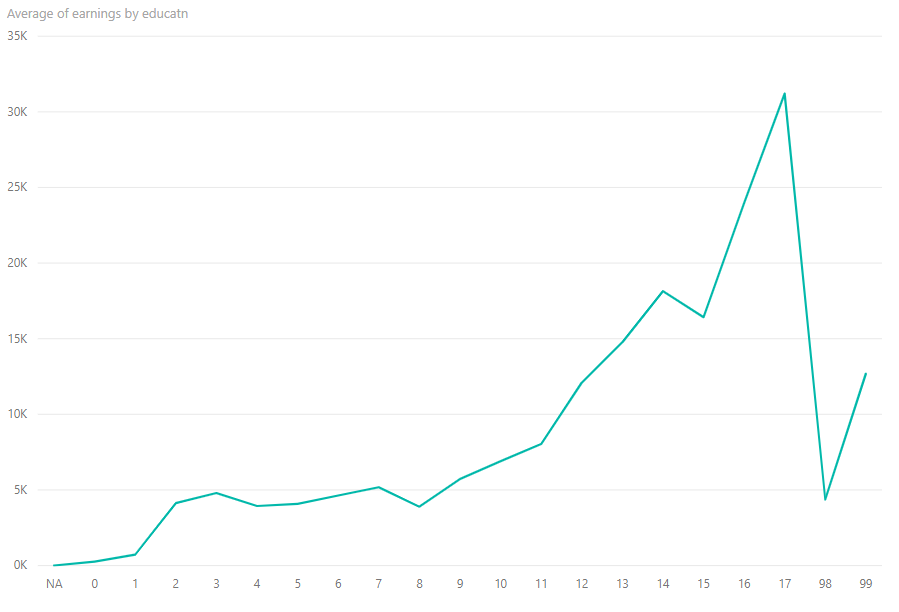
## Visualizations

**Divorced/separated stats on edu level less than 11\***



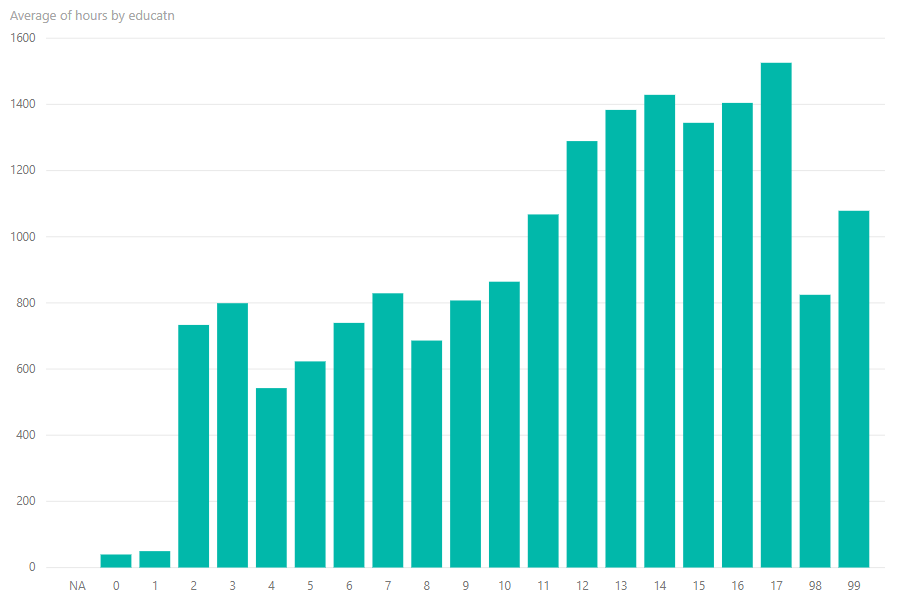
**Divorced/separated stats on edu level greater than 10**

**Average of Earning Against Education Level**



According to the graph, it provides us an indication that education level has a positive impact on the earning. When the education level increases, earning average has increased. We have considered education levels like 98, 99 are outliers.

**Average of Hours Spend Against Education Level**

****

The graph shows that with the education level, their average of working hours also has increased. This could have a negative impact for their marriage lives, since they have given more priority for work comparing to the family.