YuGu

October 1, 2019

Exercise 1

[123]:
```
import numpy as np
import pandas as pd
acs = pd.read_stata('US_ACS_2017_10pct_sample.dta')
```

Exercise 2

[121]:
```
median_inctot = acs['inctot'].median()
print(median_income)
```

```
33700.0
```

Exercise 3

[99]:
```
pd.value_counts(acs['inctot'])
```

[99]:
```
9999999    53901
0          33679
30000       4778
50000       4414
40000       4413
           ...
23260          1
187020         1
18790          1
47480          1
20470          1
Name: inctot, Length: 8471, dtype: int64
```

[100]:
```
pd.value_counts(acs['inctot'],normalize = True)
```

[100]:
```
9999999    0.168967
0          0.105575
30000      0.014978
50000      0.013837
40000      0.013834
             ...
23260      0.000003
187020     0.000003
18790      0.000003
47480      0.000003
```

```
20470      0.000003
Name: inctot, Length: 8471, dtype: float64
```

Exercise 4

[101]:
```
acs['inctot']=acs['inctot'].replace(9999999,np.nan)
```

Exercise 5

[102]:
```
np.mean(acs['inctot'])
```

[102]:
```
40890.177564946454
```

[103]:
```
a= acs[acs['inctot'].isnull()]
```

[104]:
```
a['age'].value_counts()
```

[104]:
```
10    3997
9     3977
14    3847
12    3845
13    3800
      ...
66       0
67       0
68       0
69       0
48       0
Name: age, Length: 97, dtype: int64
```

[105]:
```
a['age'].sample(20)
```

[105]:
```
83301      4
277642    14
83750     14
178139     3
318481    12
52886     11
137006    11
86427     10
81052      5
261207    11
54246      2
64584     13
262867     9
253266     4
32700     10
125891    10
242193     4
138642    14
286583    10
168152    12
Name: age, dtype: category
```

```
Categories (97, object): [less than 1 year old < 1 < 2 < 3 ... 93 < 94 < 95 <
96]
```

[106]:
```
b = acs[acs['inctot'].notnull()]
```

[107]:
```
b['age'].value_counts()
```

[107]:
```
60                      4950
54                      4821
56                      4776
59                      4776
58                      4734
                        ...
4                          0
3                          0
2                          0
1                          0
less than 1 year old       0
Name: age, Length: 97, dtype: int64
```

9999999 was used when children are counted.

Exercise 7

[108]:
```
acs['empstat'].value_counts()
```

[108]:
```
employed              148758
not in labor force    104676
n/a                    57843
unemployed              7727
Name: empstat, dtype: int64
```

[109]:
```
acs_emp=acs[acs['empstat']=='employed']
```

Exercise 8

[110]:
```
acs_emp['race'].value_counts()
```

[110]:
```
white                               116017
black/african american/negro         13175
other asian or pacific islander        6424
other race, nec                        5755
two major races                        3135
chinese                                2149
american indian or alaska native       1290
three or more major races               426
japanese                                387
Name: race, dtype: int64
```

[111]:
```
white_inctot = acs_emp[acs_emp['race']=='white']['inctot'].mean()
white_inctot
```

[111]:
```
60473.15372747098
```

[112]:
```
black_inctot = acs_emp[acs_emp['race']=='black/african american/
 ↪negro']['inctot'].mean()
black_inctot
```

[112]:
```
41747.949905123336
```

[113]:
```
x = white_inctot/black_inctot-1
print('The average White American makes {} percent more than the average Black␣
 ↪American'.format(round(x*100,2)))
```

```
The average White American makes 44.85 percent more than the average Black
American
```

More practice 1

[114]:
```
acs_emp['w_inctot'] = acs_emp['inctot']*acs_emp['perwt']
```

```
/Users/YuGu/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

[115]:
```
wwhite_inctot = acs_emp[acs_emp['race']=='white']['w_inctot'].sum()/
 ↪acs_emp['perwt'].sum()
wwhite_inctot
```

[115]:
```
43035.06956393144
```

[116]:
```
wblack_inctot = acs_emp[acs_emp['race']=='black/african american/
 ↪negro']['w_inctot'].sum()/acs_emp['perwt'].sum()
wblack_inctot
```

[116]:
```
4740.755242940651
```

[117]:
```
wwhite_nonhispanic_inctot =␣
 ↪acs_emp[(acs_emp['race']=='white')|(acs_emp['hispan']=='not␣
 ↪hispanic')]['w_inctot'].sum()/acs_emp['perwt'].sum()
wwhite_nonhispanic_inctot
```

[117]:
```
52994.82552571714
```

[118]:
```
y = wwhite_nonhispanic_inctot/wblack_inctot-1
print('The average White_nonhispanic American makes {} percent more than the␣
 ↪average Black American'.format(round(y*100,2)))
```

```
The average White_nonhispanic American makes 1017.86 percent more than the
average Black American
```