

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.formula.api as sm
import scipy.stats as st
import statistics
```

```
In [2]: df = pd.read_csv('cps_2017_cleaned.csv')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	Unnamed: 0	age	sex	uwe	loguwe	yearsEd
0	49	33	male	788.461538	6.670084	16
1	50	31	female	650.000000	6.476972	14
2	87	28	female	538.461538	6.288716	12
3	89	25	female	865.384615	6.763174	14
4	121	33	male	442.307692	6.092006	12

Part A

```
In [4]: model = sm.OLS(df['loguwe'],df['yearsEd']).fit()
model.summary()
model.params
model.bse
st.t.interval(0.95, len(df)-1, loc=np.mean(df.yearsEd), scale=st.sem(df.yearsEd))
```

```
Out[4]: (13.760913081675584, 13.897158013895512)
```

The coefficient above is .4566 which means that for every year increase in yearsEd loguwe changes by .4566%

Part B

```
In [5]: '''  
        Element structure: [coeff, se, 90%, 95%]  
        '''  
  
        final = []  
        for i in range(500):  
            element = []  
            temp = df.sample(n=100, replace=True)  
            model = sm.OLS(temp['loguwe'], temp['yearsEd']).fit()  
            element.append(float(model.params))  
            element.append(float(model.bse))  
            element.append(st.t.interval(0.90, len(temp)-1, loc=np.mean(temp.yea  
rsEd), scale=st.sem(temp.yearsEd))  
            element.append(st.t.interval(0.95, len(temp)-1, loc=np.mean(temp.yea  
rsEd), scale=st.sem(temp.yearsEd))  
            final.append(element)
```

```
In [6]: final[0]
```

```
Out[6]: [0.4571149008385476,  
         0.011885188781245776,  
         (13.248471393537141, 14.591528606462857),  
         (13.117503287370255, 14.722496712629745)]
```

Part C

```
In [7]: def find_mean(final, category):  
        total = 0  
        arr = []  
        for i in final:  
            arr.append(i[category])  
            total += i[category]  
        dev = statistics.stdev(arr)  
        return total/len(final), dev
```

```
In [8]: find_mean(final, 0)
```

```
Out[8]: (0.45727115949713376, 0.006975775046609794)
```

```
In [9]: consistency_90 = 0
        hits = 0
        for i in final:
            if .4566 in i[2]:
                hits += 1

        consistency_90 = hits/500
        consistency_90
```

Out[9]: 0.0

```
In [10]: array1 = [x[0] for x in final]
```

Part D

```
In [11]: plt.hist(array1)
```

```
Out[11]: (array([ 4., 14., 27., 51., 125., 103., 80., 66., 23., 7.]),
         array([0.43604866, 0.44002874, 0.44400882, 0.4479889 , 0.45196897,
                0.45594905, 0.45992913, 0.46390921, 0.46788929, 0.47186937,
                0.47584945]),
         <a list of 10 Patch objects>)
```

