

JianwenWu
ECO B2000
Prof. Foster

1. Group Members: Keyi Long.
2. In the dataset, the minimum wage is \$0 and the maximum wage is \$504,000. I define no college degree as people who has no high school or high school degree. Also, I define college degree as people who has some college, associate, or bachelor degree.

R code:

```
prop.table(table(cut(income_wagesal,breaks=c(- 1,75000,100000,300000,1000000)),educ_indx))
```

R-output (Unconditional)

Wages	No Hs	HS	Some Coll	AS	Bach	Adv Deg
Less than 75000	0.315622	0.213271	0.149724	0.060846	0.110573	0.071528
75000<x<100000	0.000596	0.004386	0.004814	0.003683	0.012811	0.012735
100000<x<300000	0.000346	0.001783	0.002822	0.001732	0.012439	0.015317
300000<x<1000000	0.000051	0.000153	0.000194	0.000153	0.001549	0.002873

- The proportion of people with less than a college degree earns less than \$75000 is

$$Prob. (less\ than \$75000 | less\ than\ a\ college\ degree) = \frac{0.528892}{0.536207} = 0.986$$

- The proportion of people with less than a college degree earns more than \$75, 000 is 0.0136.
- The proportion of people with college degree earn less than \$75000 is 0.888
- The proportion of people with college degree earn more than \$75000 is 0.112

Explanation: About 98.6% of people less than college earns less than \$75000, and about 1.36% of people less than college degree and earns more than \$75000. About 88.8% of people with college degree earn less than \$75000, and only 11.2% earns more than \$75000.

**R output for the proportion of people's incomes who live in NYC at age 30 or younger.
(Unconditional)**

Wages	No Hs	HS	Some Coll	AS	Bach	Adv Deg
Less than 75000	8.30E-02	1.43E-02	1.93E-02	4.15E-03	1.45E-02	4.74E-03
75000<x<100000	1.53E-05	1.12E-04	1.43E-04	7.64E-05	1.25E-03	4.64E-04
100000<x<300000	1.02E-05	4.08E-05	5.60E-05	2.55E-05	5.91E-04	5.14E-04
300000<x<1000000	0.00E+00	5.09E-06	0.00E+00	0.00E+00	5.60E-05	4.58E-05

- The proportion of people with less than a college degree at age 30 or younger earn less than \$75,000 is $(8.30E-02 + 1.43E-02)/(9.75E-02) = 0.998$. Therefore, about 99.8% of those people earn less than \$75,000. Only 0.2% of those people earn more than \$75,000.
- The proportion of people with college degree at age 30 or younger earn less than \$75000 is $(1.43E-02 + 1.91E-02 + 4.15E-03)/0.040180527 = 0.945$. Therefore, about 94.5% of those people earn less than \$7,5000. Only 5.5% of those people earn more than \$75000.,

**R output for the proportion of people's incomes who live in NYC at age 40 or older.
(Unconditional)**

Wages	No Hs	HS	Some Coll	AS	Bach	Adv Deg
Less than 75000	4.07E-02	4.46E-02	2.25E-02	9.20E-03	2.11E-02	1.42E-02
75000<x<100000	2.24E-04	1.12E-03	1.29E-03	7.44E-04	2.69E-03	2.82E-03
100000<x<300000	1.38E-04	3.62E-04	4.89E-04	2.80E-04	2.51E-03	3.37E-03
300000<x<1000000	2.55E-05	4.08E-05	3.57E-05	4.08E-05	3.92E-04	8.15E-04

- Of people who has no college degree and 40 or older, 97.8% have wage of 75,000 or less than \$75,000. Only 2.2% have wage of \$75,000 or more than \$75,000.
- Of people who has college degree and 40 or older, 86.2% have wage of \$75,000 or less than \$75,000. Only 13.8% have wage of \$75,000 or more than \$75,000.

Overall conclusion: According to the analysis of income wages of NYC above, people with college degree earn more than people with no college degree. Also, people who are older earn more than people who are younger.

3. We have found some additional data for a Benford's Law analysis through Yahoo.Finance and we picked up some company's historic data imported in R. First of all, we choose a company, "FXCM" and downloaded 5 years daily stock price. We got results of "Volume" and followed by below commands in R:

```
> attach(table)

> bfd.cp <-(Volume)

> bfd.cp <- benford(Volume,1,"positive", TRUE,3)

> bfd.cp
```

Benford object:

```
Data: Volume
Number of observations used = 1258
Number of obs. for second order = 786
First digits analysed = 1
```

Mantissa:

Statistic	Value
Mean	0.488
Var	0.079
Ex.Kurtosis	-1.129
Skewness	0.034

The 5 largest deviations:

digits	absolute.diff
1	23.48
2	20.95
3	18.83
4	14.91
5	7.56

Stats:

Pearson's Chi-squared test

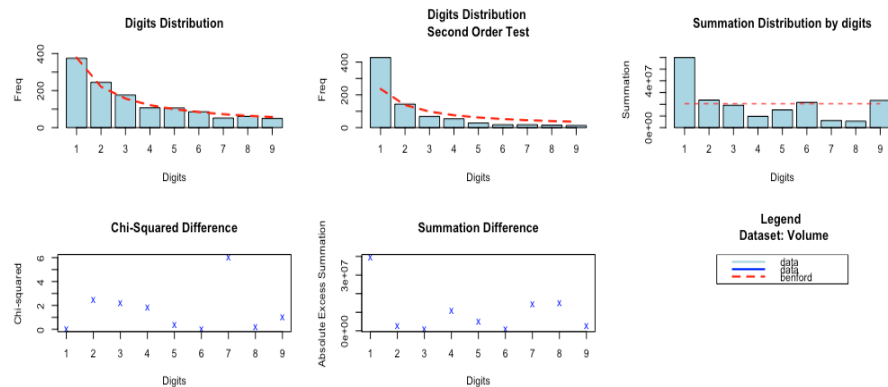
```
data: Volume
X-squared = 14.238, df = 8, p-value = 0.07578
```

Mantissa Arc Test

```
data: Volume
L2 = 0.0020198, df = 2, p-value = 0.07879
```

```
Mean Absolute Deviation: 0.008916304
Distortion Factor: -3.452861
```

Remember: Real data will never conform perfectly to Benford's Law. You should not focus on p-values!



According to the graphs, the historic “Volume” variable is closely follow Benford’s Law, and when we choose other observed selected variable, such as “Close”, we also got plot which is close to follow the Law as well.

4. **Statistics:**

- The average income salary in NYC is approximate \$22727.42.
- The average income salary in Brooklyn, Manhattan, Staten island, and Bronx are \$20323.16, \$44704.39, \$23913.52, \$14022.08, and \$20256.13 respectively.
- The average owner cost in Brooklyn, Manhattan, Staten island, and Bronx are \$ 832.3255, \$579.2995, \$1558.796, 453.1302, and \$1122.255 respectively.

Explanation: According to my data above, People that in Manhattan have the highest income salary, and they have second lowest owner cost. Based on that, there would be a huge gap. Also, based on data in #1, the people who have higher educational background earn the most.

Reference:

- FXCM Historical Prices | FXCM Inc. Class A Common Stock Stock - Yahoo Finance By In - <http://finance.yahoo.com/quote/FXCM/history?period1=1315800000>
- Cinelli, Carlos. "Package ‘benford.analysis’." <https://cran.r-project.org/web/packages/benford.analysis/benford.analysis.pdf>. N.p., 29 Aug. 2016. Web. 12 Sept. 2016.