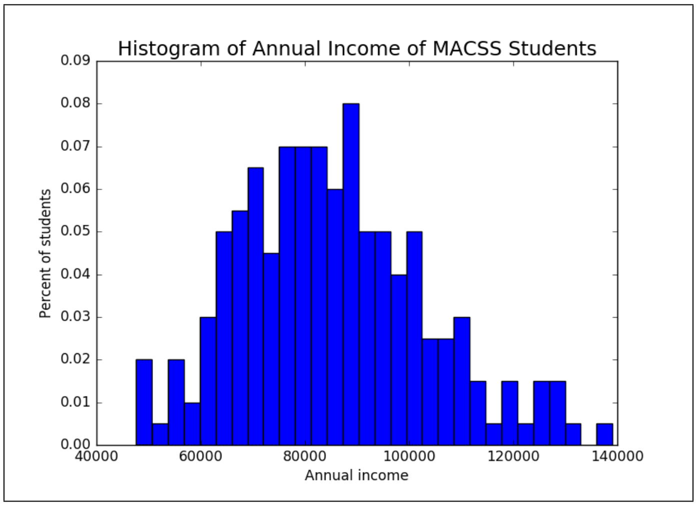
Problem Set #2

MACS 30100, Dr. Evans
Xinzhu Sun(12147991)

Problem 1

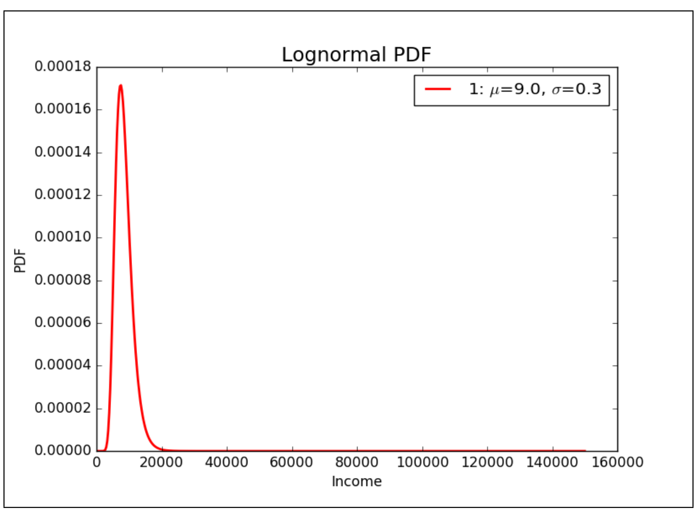
Part (a). Plot a histogram of percentages of the income

The histogram of annual incomes of students who graduated in 2018, 2019, and 2020 from the University of Chicago M.A. Program in Computational Social Science is as following:

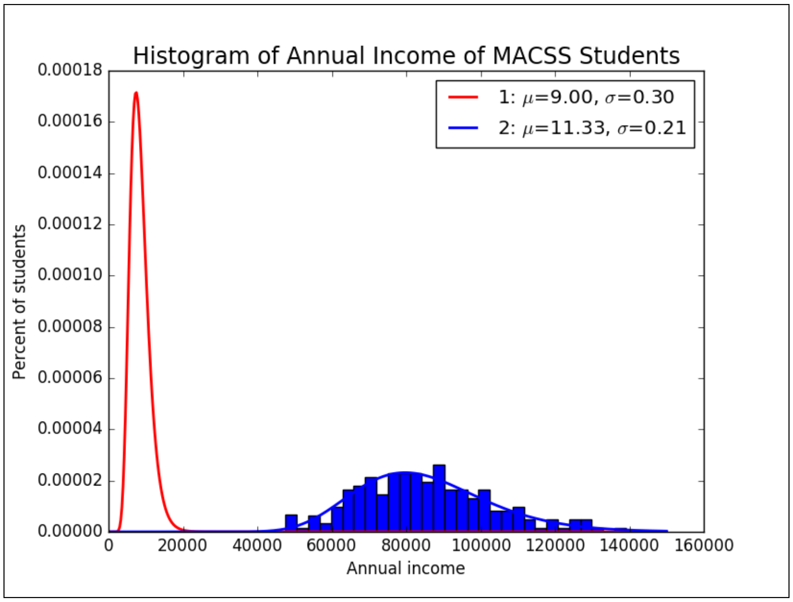


Part (b).Plot the lognormal PDF

The lognormal pdf with $\mu = 9.0$ and $\sigma = 0.3$, for $0 \leq x \leq 150{,}000$ is as following. The log-likelihood value is -8298.636956005032.



Part (c). Estimate the lognormal distribution by ML



The MLE parameters of lognormal distribution are: $\mu = 11.3314403048$, σ= 0.211674604497.The value of likelihood function is -2239.5347439980105. The variance-covariance matrix is:

$$VCV_{(MLE)} = \begin{bmatrix} 0.05134753 & 0.01869686 \\ 0.01869686 & 0.00694917 \end{bmatrix}$$

Part (d). Perform chi squared test

χ^2 of H_0 with 2 degrees of freedom p-value = 0.0 .
That is, the likelihood that the income data came from the distribution is very low. We should reject the null hypothesis that μ is 9.0 and σ is 0.3.

Part (e). Report the probabilities using the estimated model

The possibility of having an income higher than \$100000 is 19.58% . The possibility of having an income lower than \$75000 is 30.77% .

Problem 2

Part (a). Estimate the parameters of the linear mode and report

The estimated parameters are as following:

$$\beta_0 = 0.25164631958$$
$$\beta_1 = 0.012933347853$$
$$\beta_2 = 0.400502072911$$
$$\beta_3 = -0.00999167071918$$
$$\sigma_{MLE} = 0.00301768213759$$

The maximized log-likelihood is 876.8650464331619. The estimated variance covariance matrix of the estimates is:

$$VCV_{(MLE)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Part (b). Use chi squared test to determine the probabilities

χ^2 of H_0 with 5 degrees of freedom p-value = 0.0 .

That is, the likelihood that age, number of children, and average winter temperature have no effect on the sick days is very low. We should reject the null hypothesis that σ^2 is 0.01 and β_1,β_2,$\beta_3 = 0$.