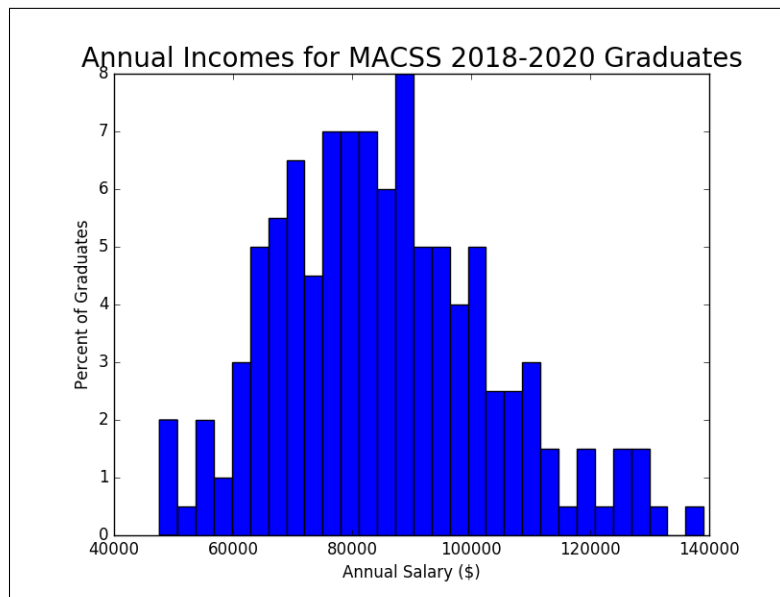


Problem Set #[2]
MACS 30100, Dr. Evans
Joanna Tung

Problem 1

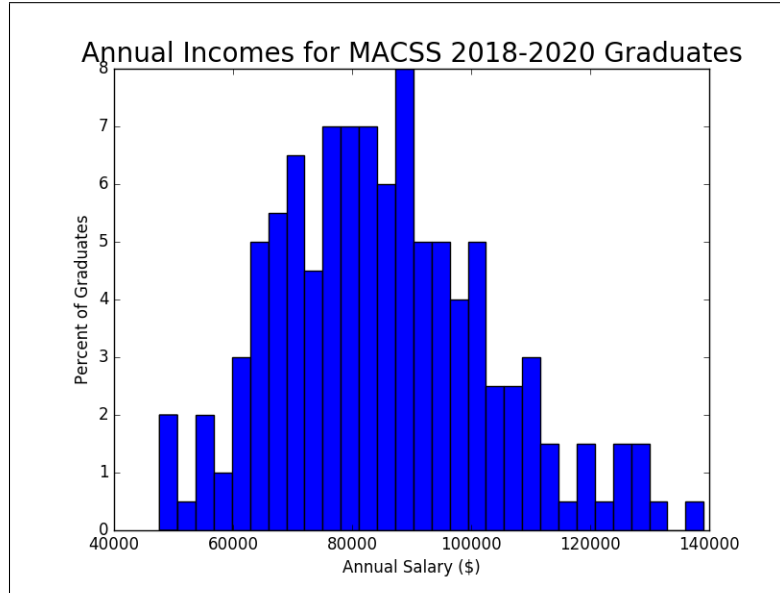
Part 1a. Below is the histogram illustrating the percent of MACS 2018-2020 graduates and their annual incomes, generated from the income.txt file.

**Figure 1: Histogram of Annual Incomes for
MACS 2018-2020 Graduates**



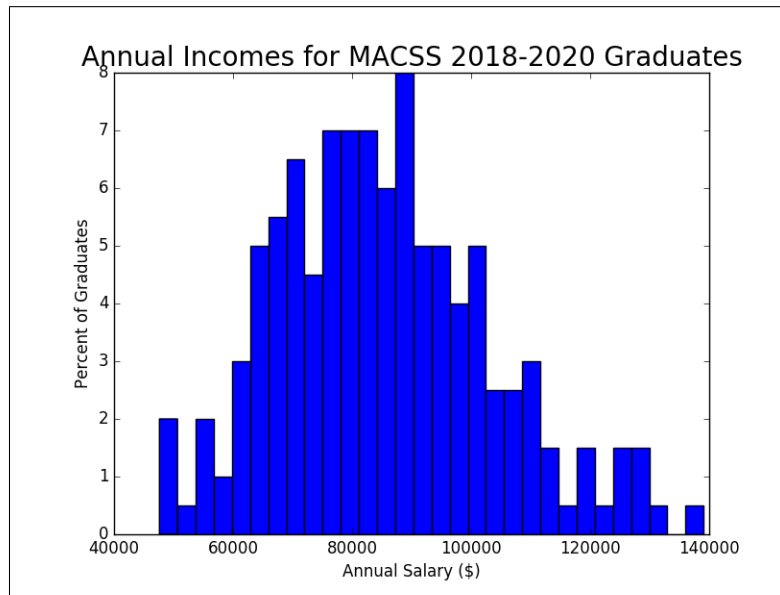
Part 1b. Below is the histogram illustrating the normalized number of MACS 2018-2020 graduates and their annual incomes, generated from the income.txt file. The lognormal probability distribution function with mean (μ) = 9 and standard deviation (σ) = 0.3 has been plotted in red. This red line represents the initial "guess" for the μ and σ parameters for our model (in this problem, the model used is the lognormal pdf).

Figure 2: Histogram of Annual Incomes for MACS 2018-2020 Graduates, Normalized



Part 1c. The initial "guess" for the parameters μ and σ were used to estimate the true values of these same parameters using maximum likelihood method. Optimization methods 'TNC', 'L-BFGS-B' and 'SLSQP' were all attempted. Method SLSQP returned what seemed to be the best fit and was the only method to successfully terminate; however, no variance-covariance matrix was returned. Only method TNC returned the inverse hessian. The value of the parameters μ and σ are reported below for both method SLSQP and TNC. The variance-covariance matrix (inverse hessian) has only been provided for method TNC. The estimated parameter values for μ and σ as returned by the MLE method SLSQP has been plotted in green in the histogram below. The value of the log-likelihood function using MLE method SLSQP is -2239.53474401.

Figure 3: Histogram of Annual Incomes for MACS 2018-2020 Graduates, Normalized



Method	Item	Value
SLSQP	μ	11.3314417737
SLSQP	σ	0.211676166182
TNC	μ	11.2887517273
SLSQP	σ	0.681189027336

1.0	0
0	1.0

Part 1d. The likelihood ration test was performed using outputs from MLE method SLSQP. The chi-squared of H0 with 2 degrees of freedom p-value is 0.0. This tells us that it is unlikely that the data in incomes.txt came from the initial distribution which we "guessed" in part b.

Part 1e. Using the parameters obtained from MLE method SLSQP in our model, the probability that I will earn greater than \$100,000 per year after graduation is 19.58%. The probability that I will earn less than \$75,000 per year after graduation is 30.77%.

Problem 2

Part 2a. An initial guess of parameters (taken from part b) $\beta_0 = 1.0$, $\beta_1, \beta_2, \beta_3 = 0$, and $\sigma = 0.1$ were used to maximize the likelihood of seeing the data in datafile sick.txt. Optimization methods 'TNC', 'L-BFGS-B' and 'SLSQP' were all attempted. Only method SLSQP terminated successfully, while only Method L-BFGS-B returned the

inverse Hessian. The parameter outputs from method SLSQP were input as the initial parameters for Method L-BFGS-B to attempt to improve the optimization. The initial parameters used with Method L-BFGS-B (that is, the outputs from method SLSQP) are reported below, alongside the outputs from the optimization (parameter estimates, variance-covariance matrix). The value of the log-likelihood function for this optimization procedure is 0.

Parameter	Initial Parameter Estimate	Output MLE Parameter estimate
β_0	46.0198488659	46.01984887
β_1	103.89834802	103.89834802
β_2	-82.603516994	-82.60351699
β_3	-83.0455650978	-83.0455651
σ	-420.918478405	-420.91847841

1.0	0	0	0	0
0	1.0	0	0	0
0	0	1.0	0	0
0	0	0	1.0	0
0	0	0	0	1.0

Part 2b. A likelihood ratio test was used to determine the probability that our initial guess estimates $\beta_0 = 1.0$, $\beta_1, \beta_2, \beta_3 = 0$, and $\sigma = 0.1$ fit the model we created using MLE method L-BFGS-B. The chi-squared of H0 with 5 degrees of freedom p-value is 0.0. This tells us that it is unlikely that these parameters are correct: age, number of children, and average winter temperature should have an effect on the number of sick days.