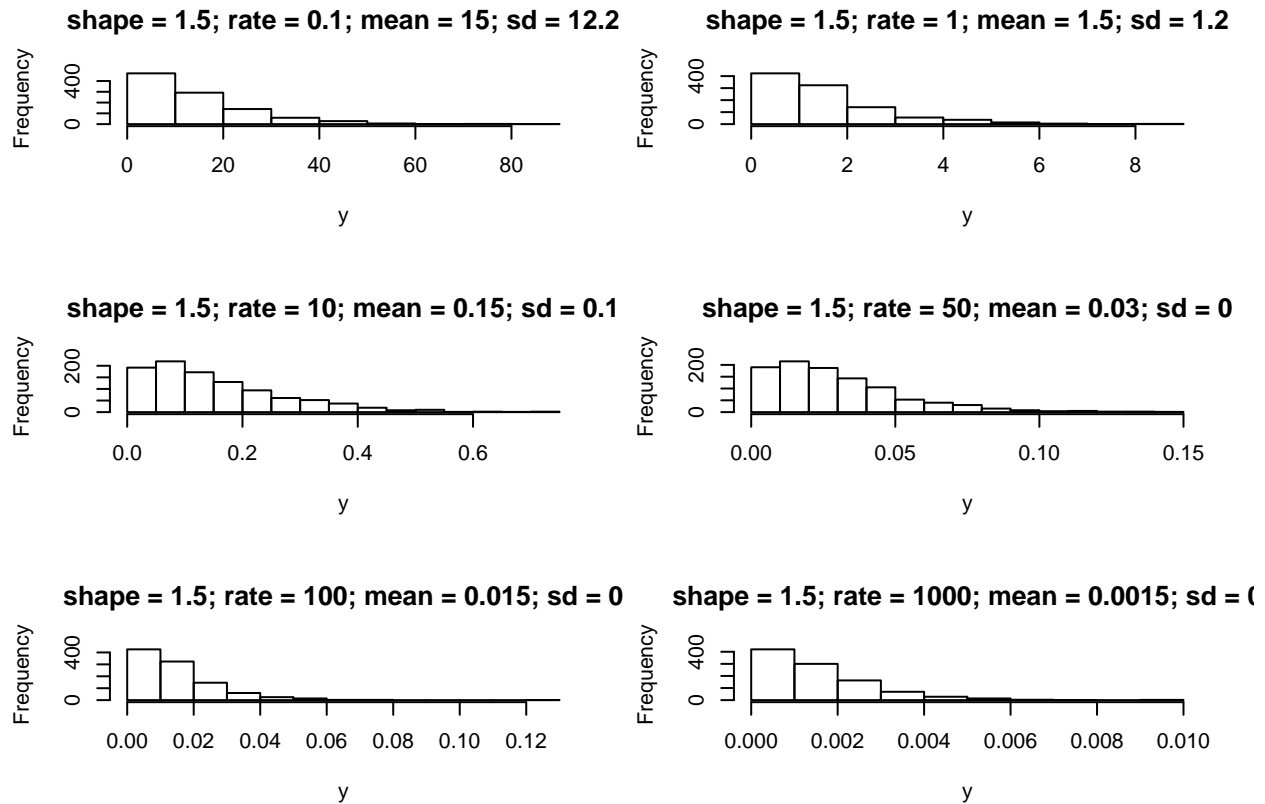# NR913 HW01

Sophie Wulfing

2/4/2022

Question 1

Explore what happens when changing the rate of a gamma distribution while holding the shape constant, similar to the "visualizing distributions" exercise from class. Provide a 6-panel figure with the 6 histograms that result. Interpret the figure.


Here you can see that by holding the shape, the data will remain concentrated in (generally) the same area. However, the right "tail" of the data is greatly affected by the rate as smaller rates will result in longer tails, as well as means that are further from zero. As the mean of the data is shape/rate, this makes sense that as the rate approaches infinity, the mean will approach zero, regardless of the shape. Further, standard deviations also decrease with higher rates (the last three graphs say SD = 0 but this is just where I chose to round) which further indicates more concentrated data at higher rates. So that's cool.

Question 2

Show how the sampling distribution of the mean is normally distributed for a distribution of sampled data that we did not cover in class, e.g., the uniform, the lognormal, beta, etc. To show this, provide a 2-panel figure with 2 histograms: one for the raw (simulated) data, and one for the sampling distribution of the means.

```
## [1] 0.05314645
```

Entire population – lognormal

Sample means – lognormal
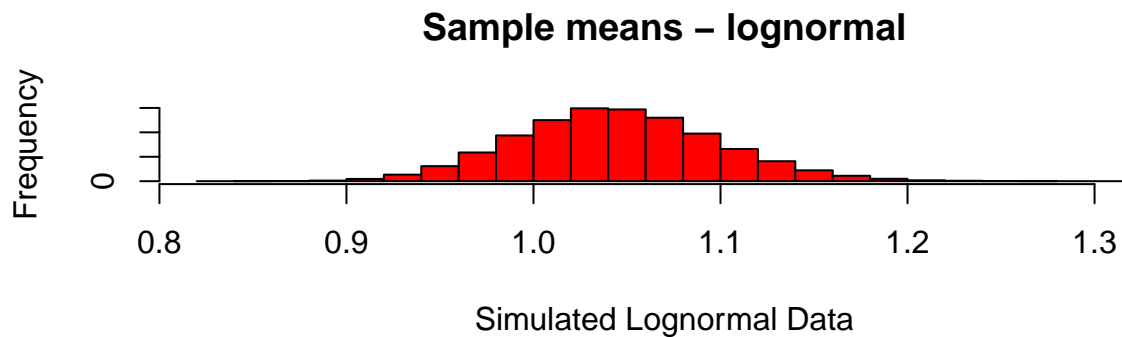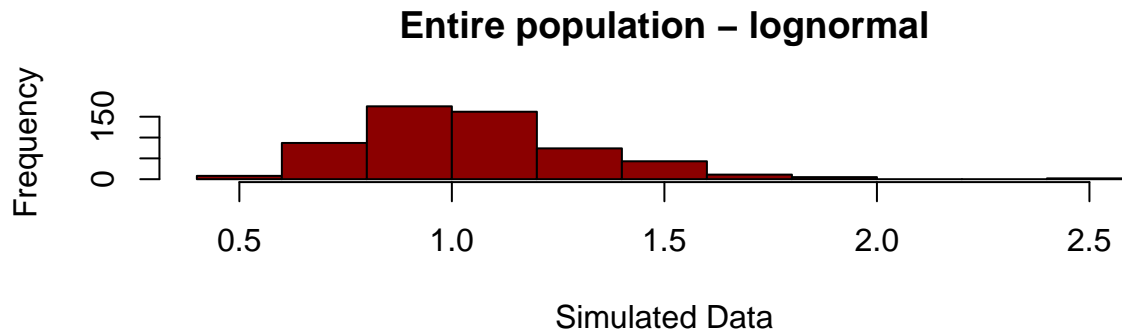
Question 4

Simulate a dataset similar to the one in the "simple linear model" exercise we completed in class. Fit a linear model to it. Provide a summary of the linear model as a table similar to that below. Provide a figure showing the model-fit prediction line, and the true relationship. Interpret your findings.

```
## $coefficients
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 14.932464 0.35662094   41.87209 2.901982e-165
## x_1          1.758284 0.01217709  144.39274  0.000000e+00


## $coefficients
##               Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 13.296790  4.4461762   2.990612 2.922043e-03
## x_2          1.873911  0.1499869  12.493829 2.381869e-31
```

a)   **b)**
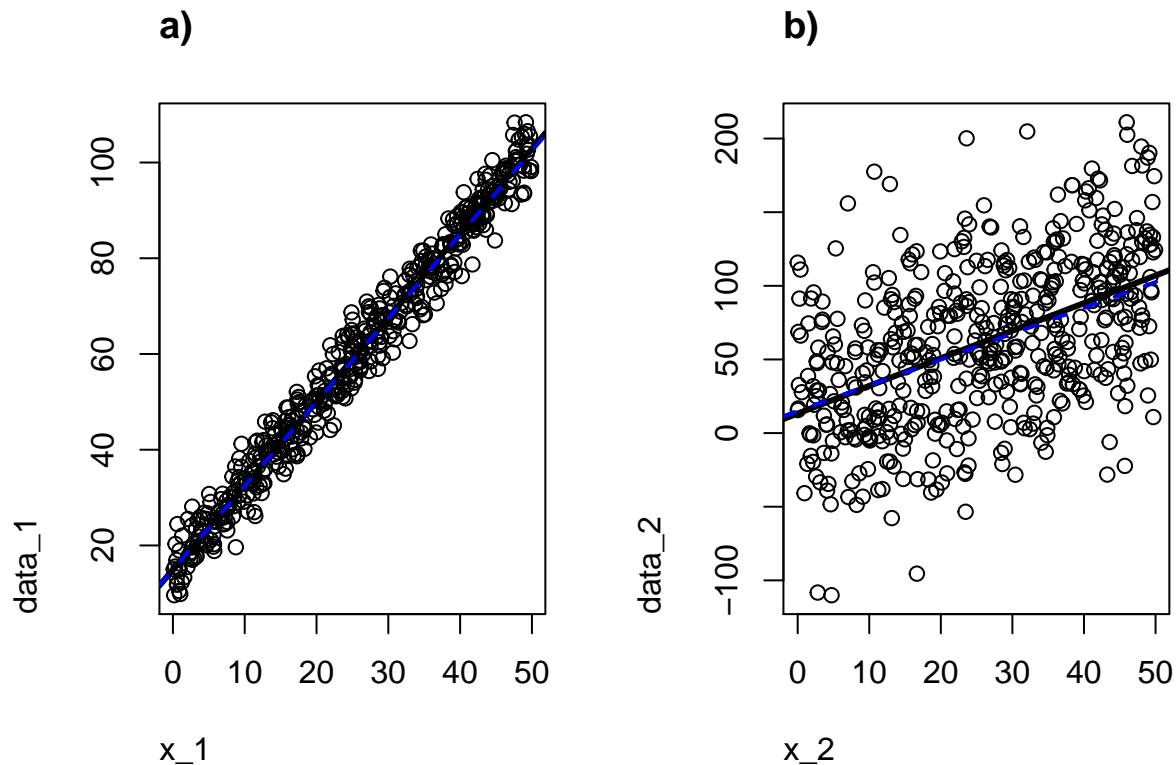
I was interested in seeing how good our linear model was at predicting b0 and b1 at different standard deviations (SD = 4 and SD = 50) of normally distributed data. Both gave very accurate predictions (both had a p value close to zero) however, the SD = 4 graph gave p values significantly closer to zero as a smaller spread of data would result in more tightly bounded datapoints.