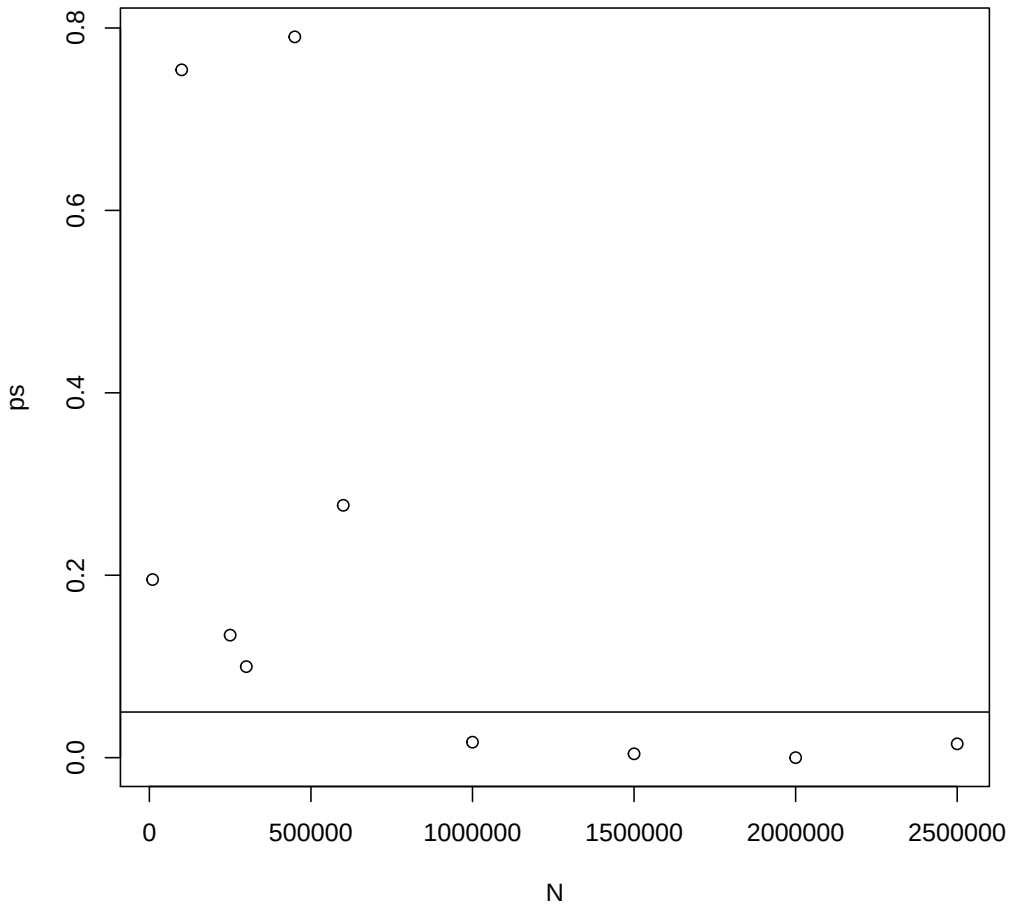
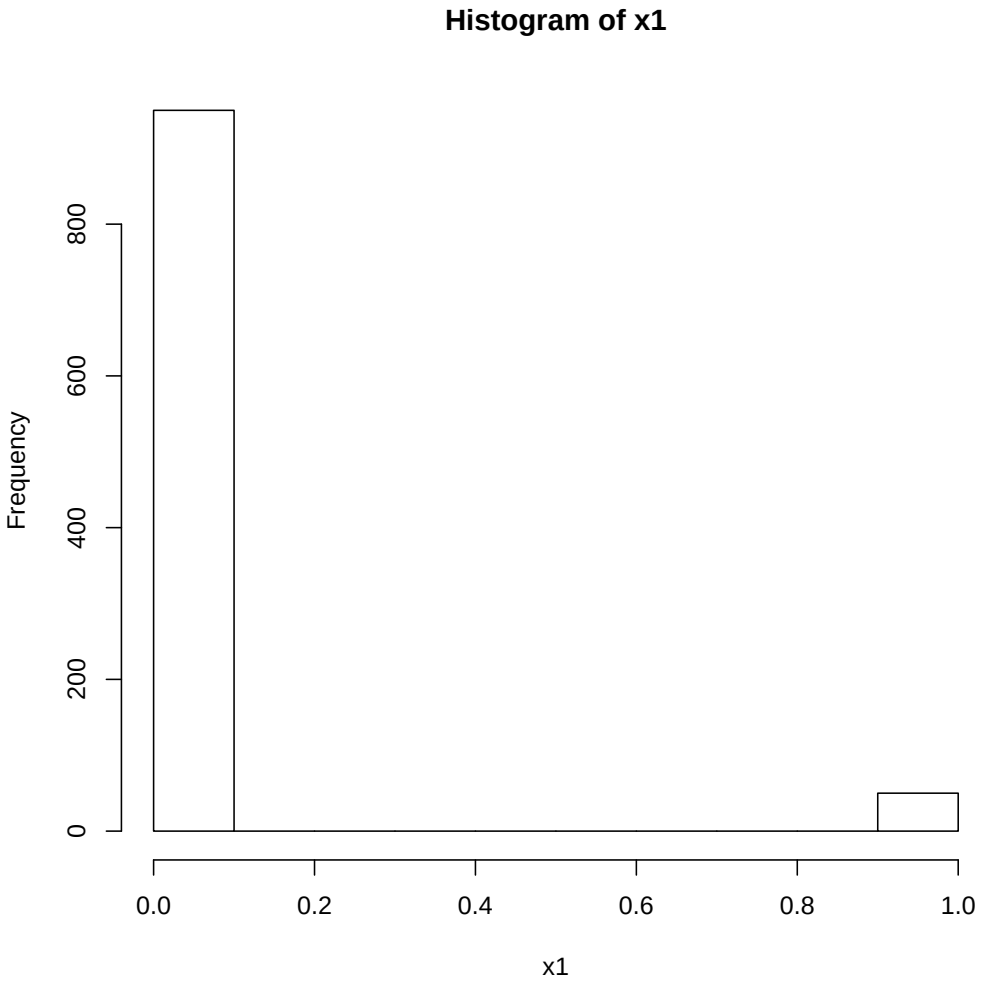
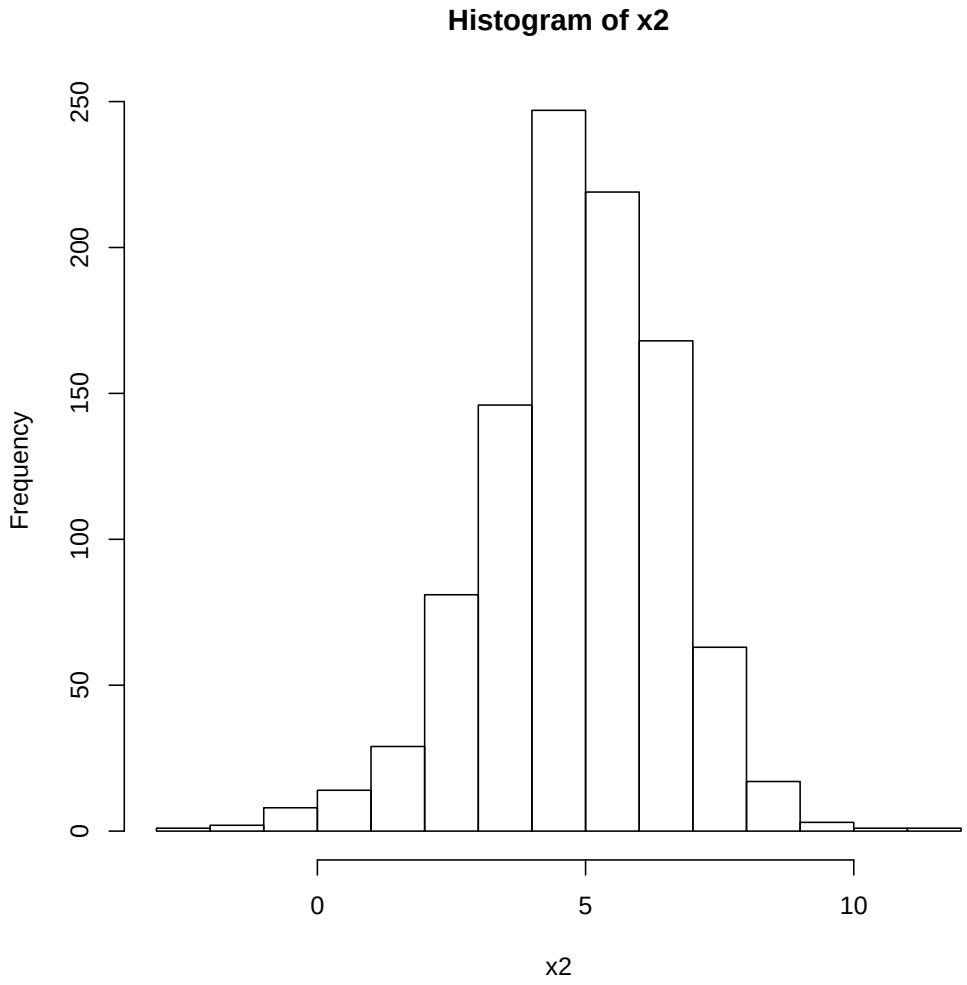
Problem 1

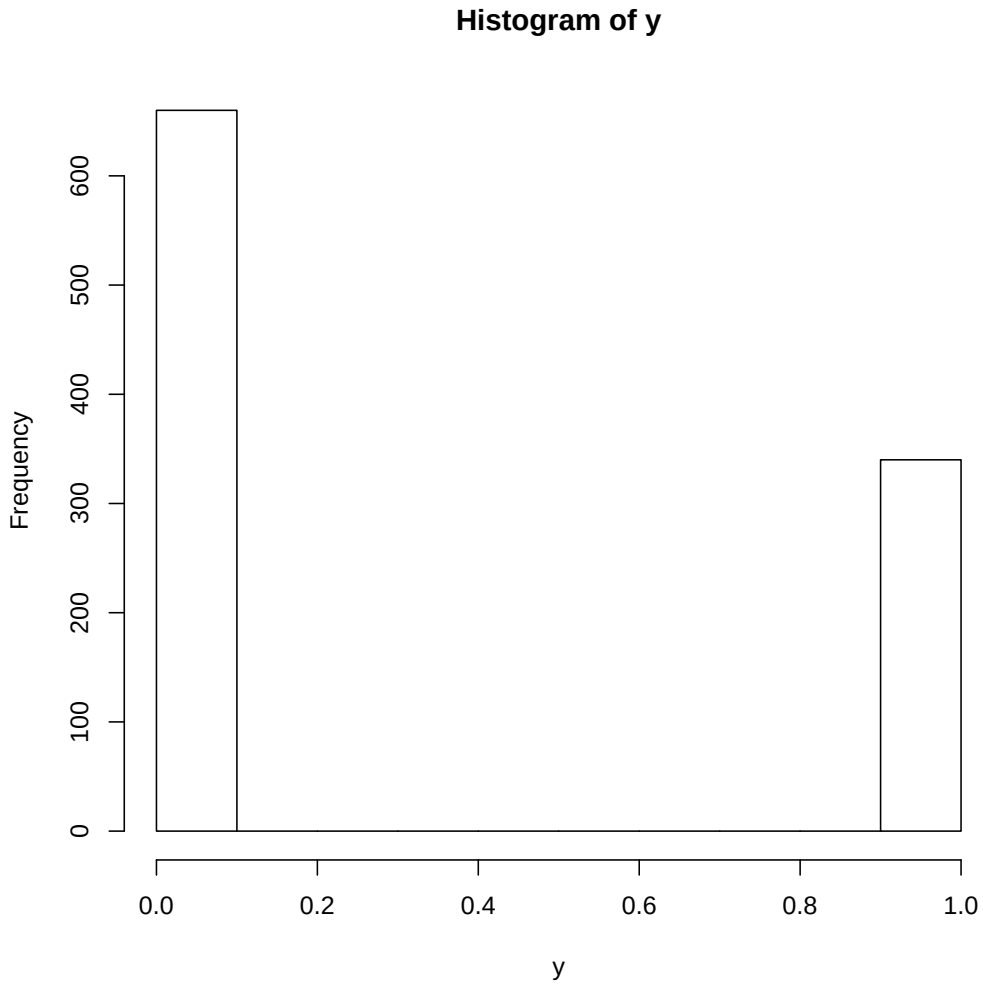


We set the parameters to $y = 3 + 0.01 * x + e$, for which y has a weak relationship with x. The correlation is 0.002. We then ran regressions on samples of increasing size, from 10000 to 2500000 and plotted the p-values for each regression. As seen in the figure, the relationship is consistently statistically significant after 1000000 samples, which shows that as sample size increases, even weak relationships will be shown as significant.

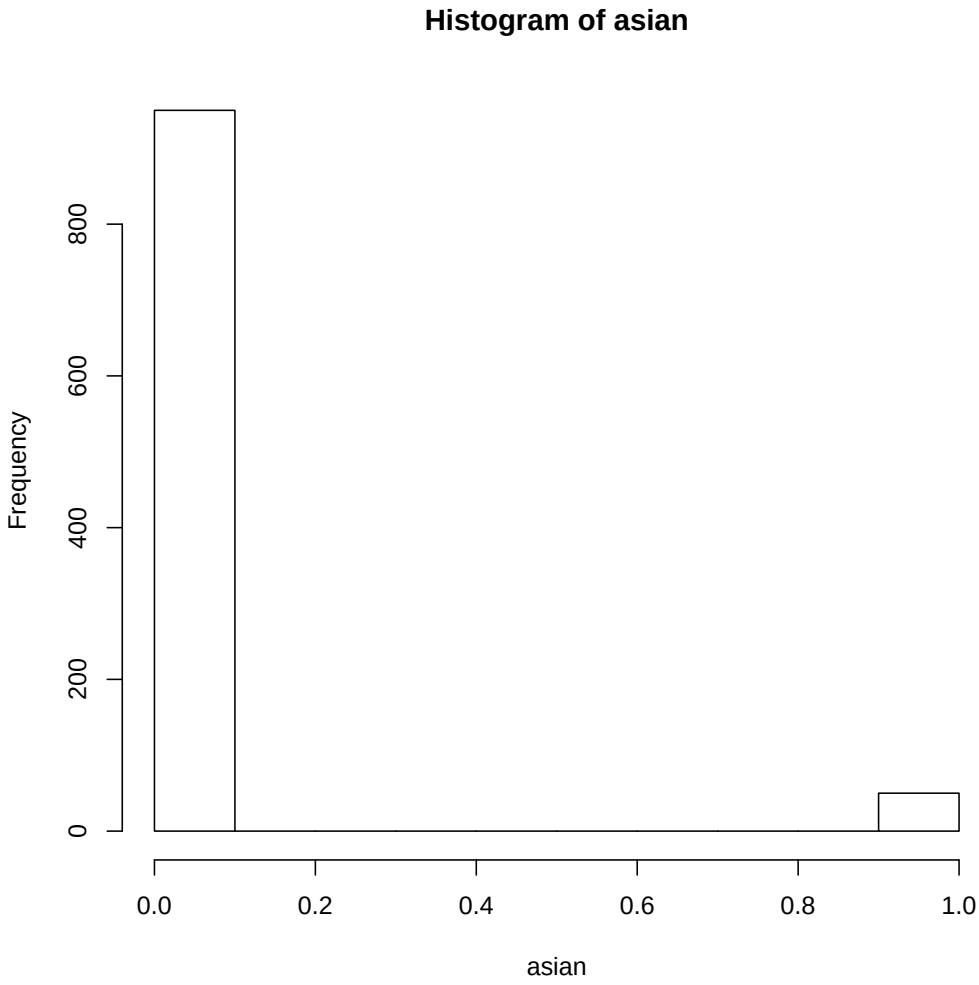
Problem 2

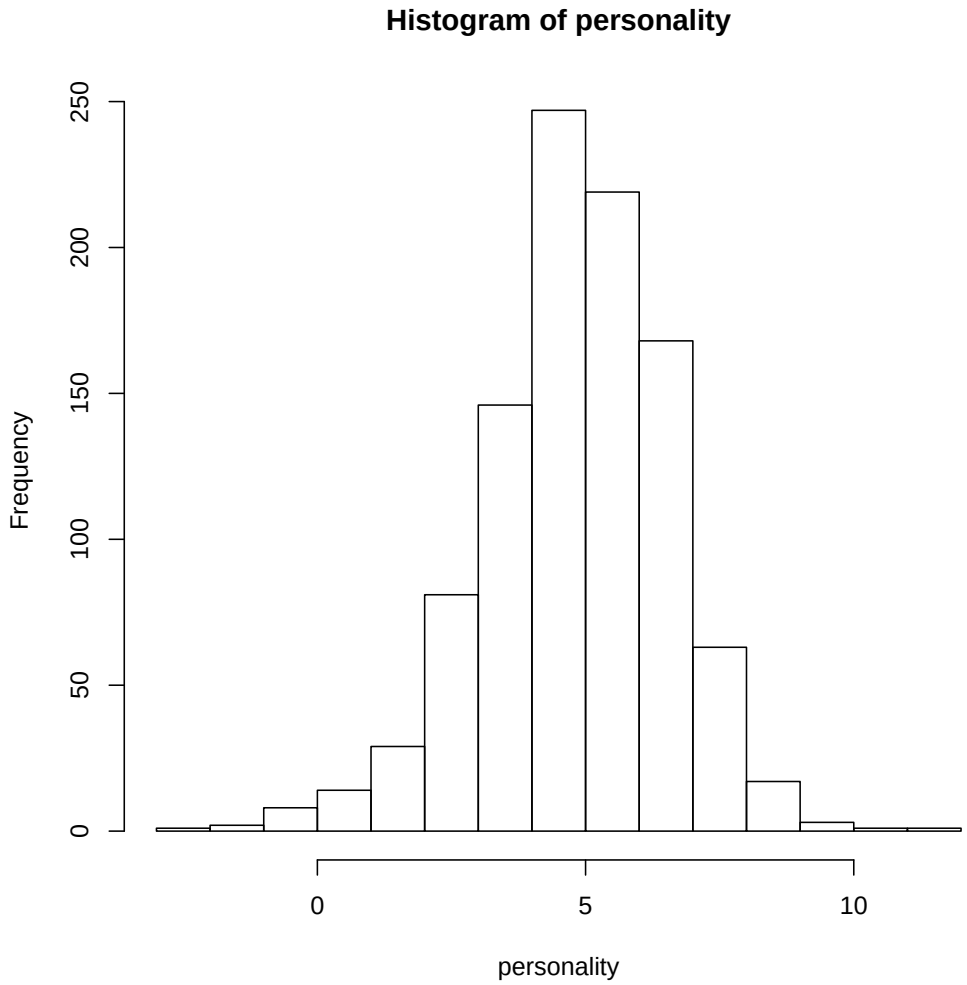

Histogram of x1
Frequency
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0
x1

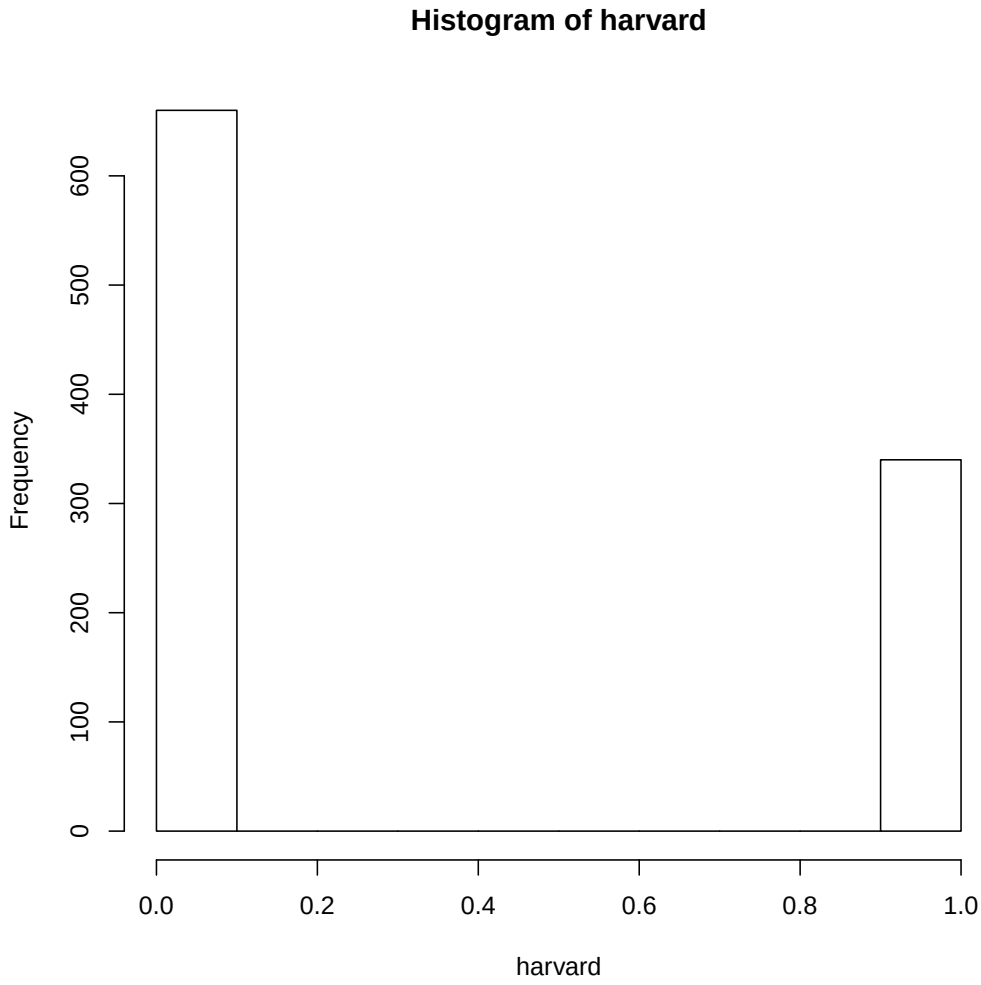

Histogram of x2
Frequency
0
50
100
150
200
250
0
5
10
x2



We completed Problem 2b prior to 2a, which is why the code is nearly identical. The correlation between x1 and y is -0.165, while the coefficient in the regression is 0.108, showing that the lurking variable x2 changes the sign.


Histogram of asian
Frequency
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0
asian


Histogram of personality
Frequency
0
50
100
150
200
250
0
5
10
personality

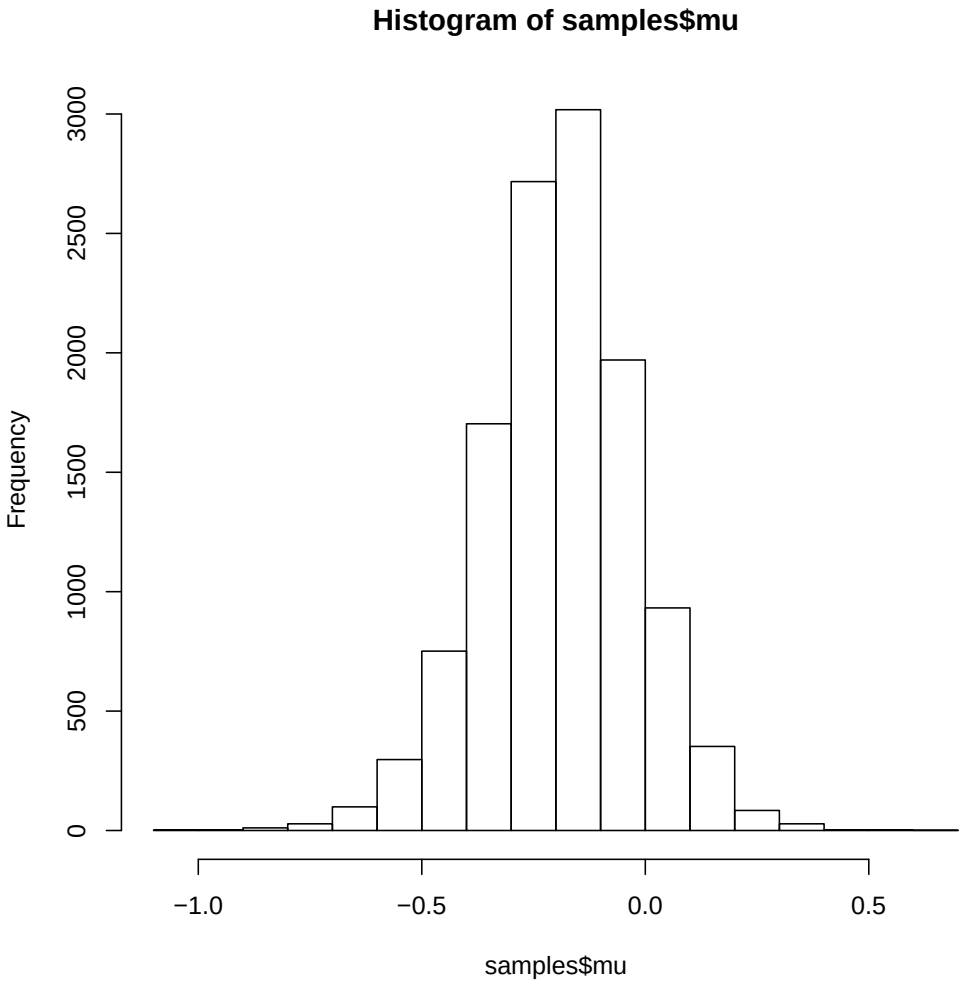


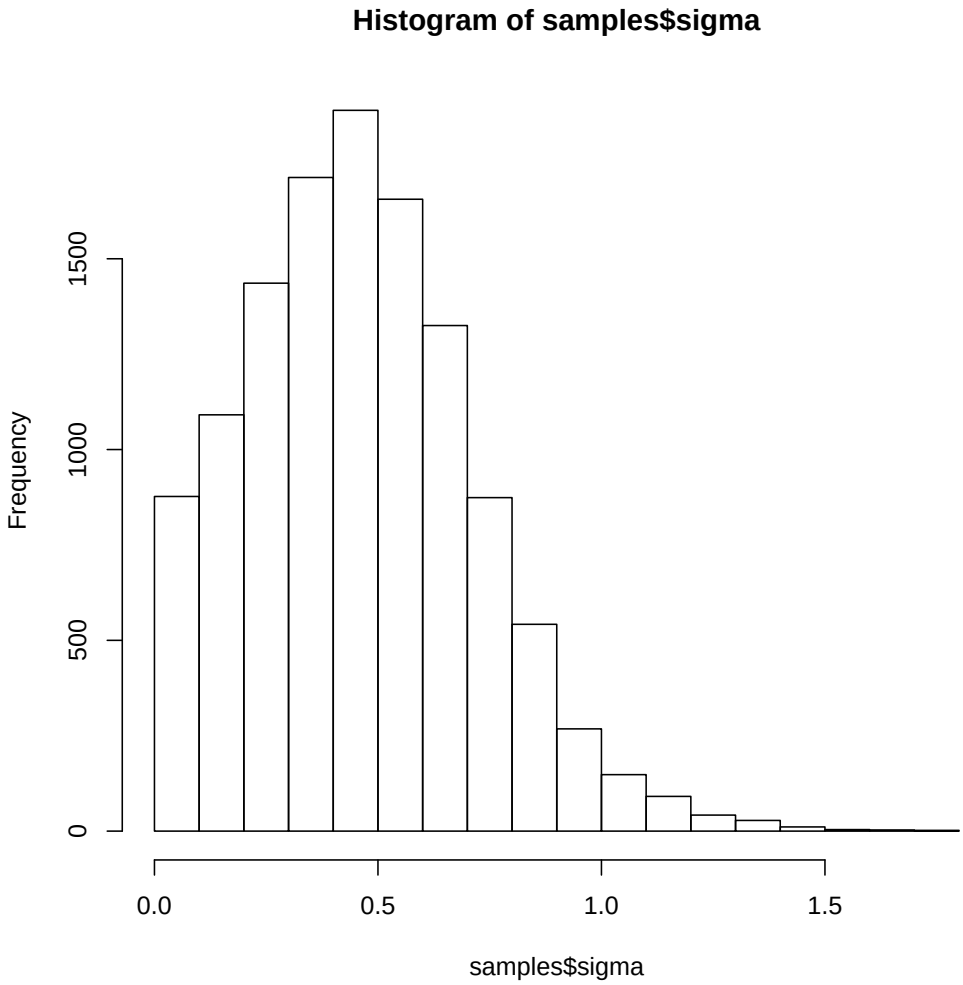
We chose to comment on the Harvard Circut Court decision on admission of Asian American applicants. Here, x1 is binomially distributed to mark if the candidate is Asian (5.6% of the U.S. population is Asian). x2 is the "personality" score as described in the brief, which here is lowered if the applicant is Asian. y is binomially distributed for acceptance to Harvard based on the personality score distribution, which is lower for Asian applicants.

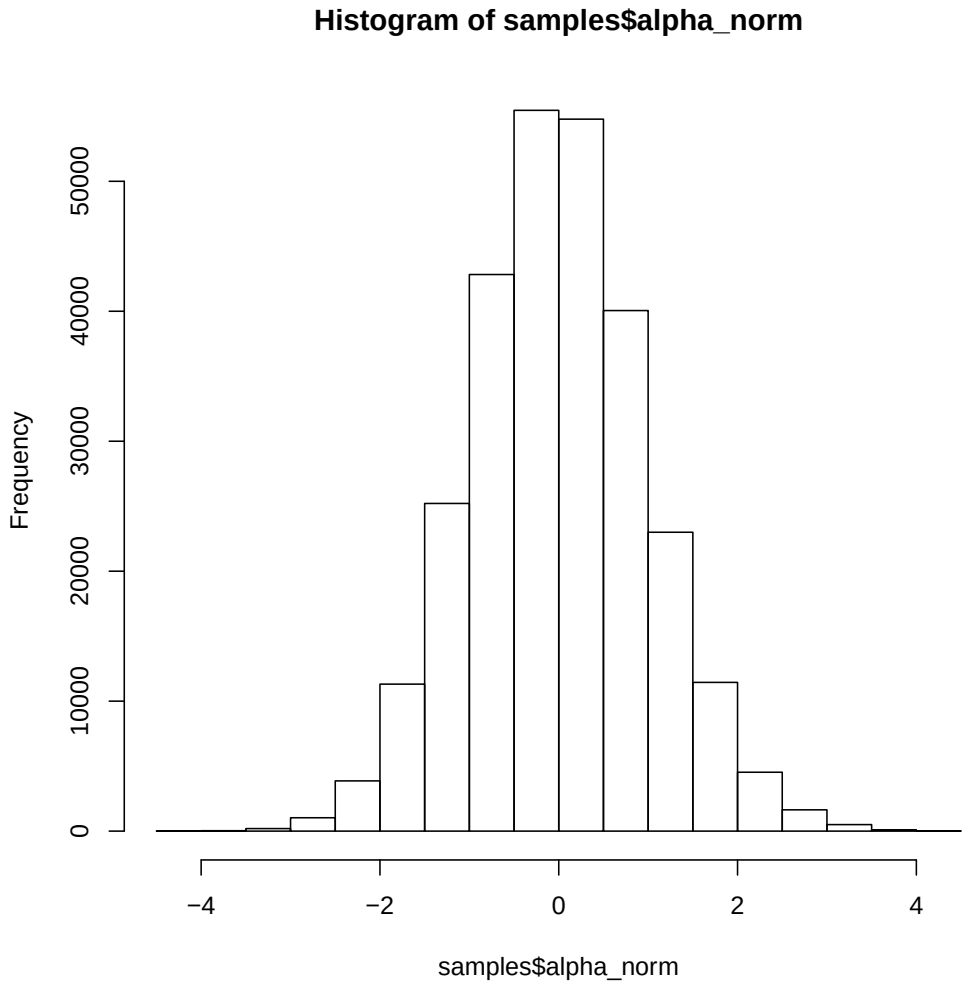
Problem 3

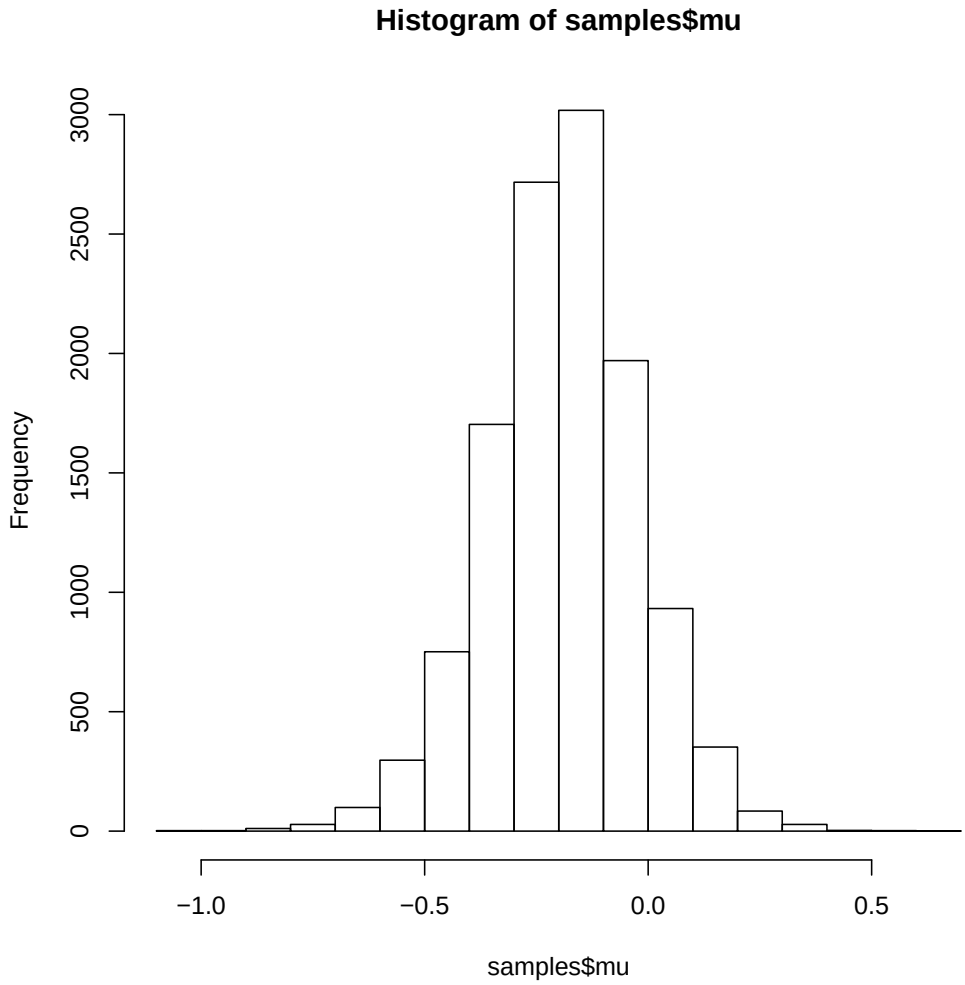
Part a

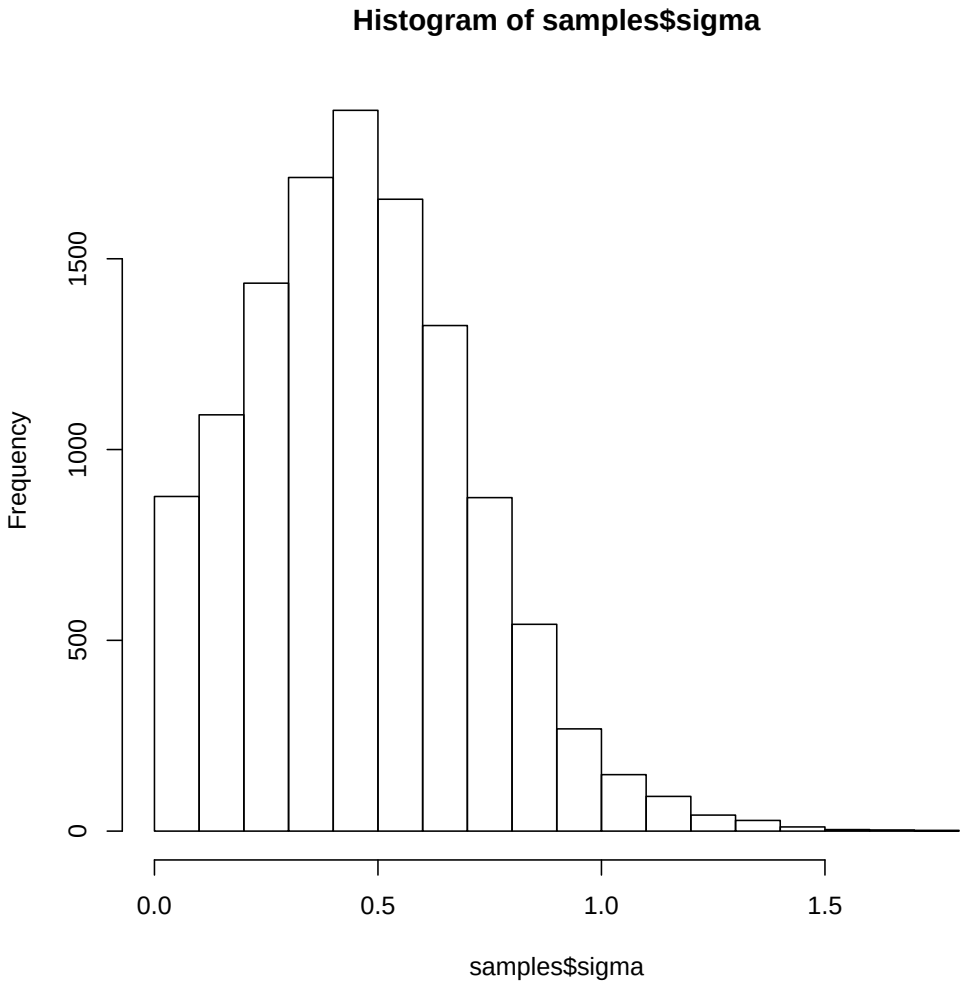
Part b

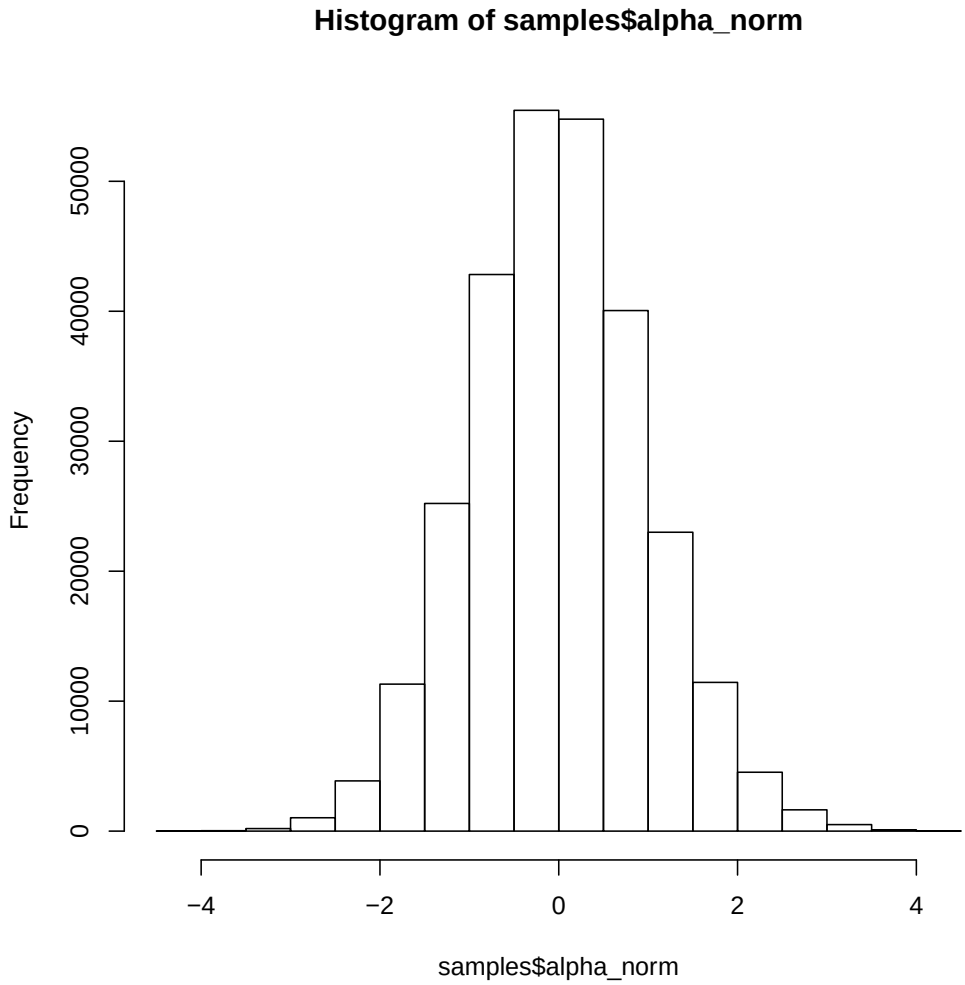

Histogram of samples$mu
Frequency
0
500
1000
1500
2000
2500
3000
−1.0
−0.5
0.0
0.5
samples$mu


Histogram of samples$sigma
Frequency
0
500
1000
1500
0.0
0.5
1.0
1.5
samples$sigma


Histogram of samples$alpha_norm
Frequency
0
10000
20000
30000
40000
50000
−4
−2
0
2
4
samples$alpha_norm


Histogram of samples$mu
Frequency
0
500
1000
1500
2000
2500
3000
−1.0
−0.5
0.0
0.5
samples$mu


Histogram of samples$sigma
Frequency
0
500
1000
1500
0.0
0.5
1.0
1.5
samples$sigma


Histogram of samples$alpha_norm
Frequency
0
10000
20000
30000
40000
50000
−4
−2
0
2
4
samples$alpha_norm