10/10

(+) nice organization
(+) extra effort into figure formatting & labeling
(+) good descriptions & observations

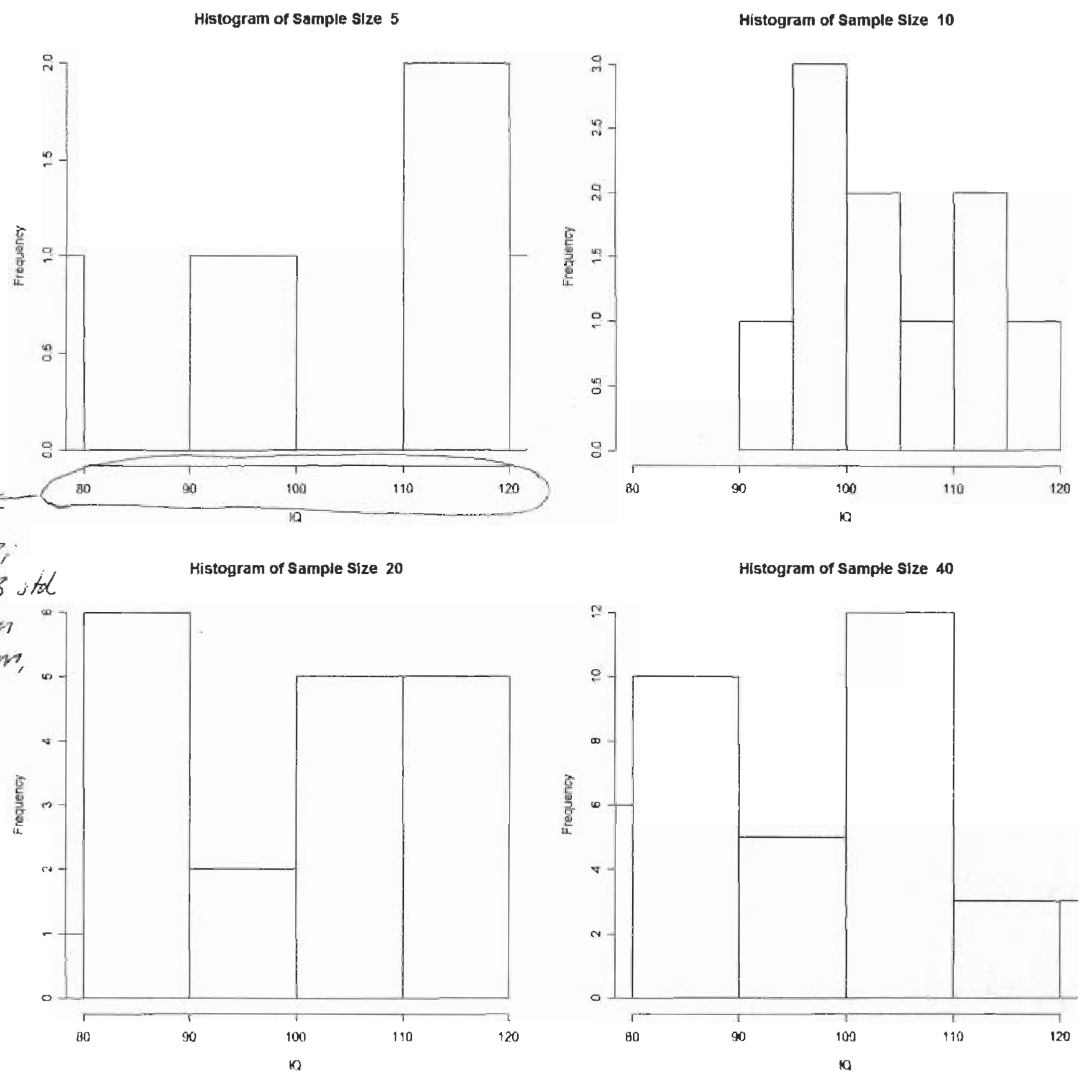
Exercise 1

(-) make sure not to cut out meaningful data

(-) occasional errors in use of terms

In this project we~~re~~ were given the task of simulating many repetitions of sampling a distribution of IQ scores with different sample sizes. We assume that the distribution is Gaussian and centered on a mean of 100 with a standard deviation of 15. Five sample sizes were used overall: 5,10,20,40, and 80. In each simulation a Gaussian distribution was sampled 1000 times with each of the sample sizes listed above. Figure 1 illustrates a histogram for one repetition of 1000 for each sample size. Histogram of a typical sample.

needs to be a wider range; you use only 1.3 std wide maximum and minimum, cutting off too much of your data



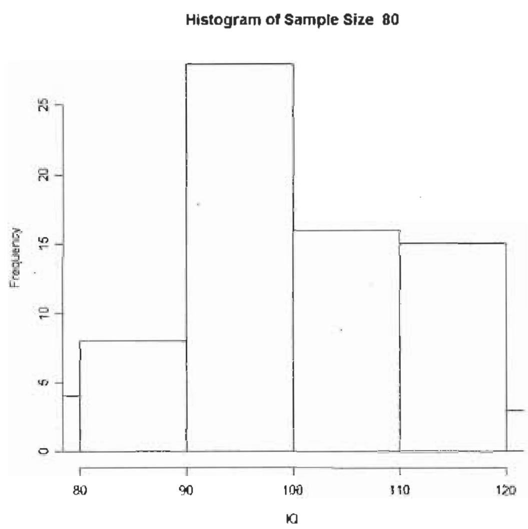


***Figure 1:** Examples of data collected from a single trial/repetition from a Gaussian distribution (μ= 100, σ= 15) with sample sizes of 5,10,20,40 and 80*

These results illustrate that with ~~more samples per trial~~ larger sample size the underlying normal distribution of the population becomes more apparent. For each sample that we draw we can calculate the mean and standard deviation and examine the properties of the sampling distribution for both statistics. Given the results above one has an intuitive feel that ~~more~~ larger samples convey more information about the population and will decrease our uncertainty. Figures 2 and 3 below show histograms of the sampling distributions of the mean and standard deviation respectively, generated for each sample size over 1000 repetitions. With increasing sample size the variance of both sampling distributions decreases yielding sharper peaks in these sampling distribution as sample size increases. This means that with increased sample size our estimate of the mean and standard deviation are reduced. Formally this result is captured as the standard error of the mean (SEM) and allows one to place confidence intervals on the probability that we would obtain a mean from a given sample size. The formula for the SEM is:

our uncertainty about those estimates is reduced

obtain a certainly precise approximation of the mean

$$SEM = SD_{sample}/\sqrt{n}$$

Figure 4 plots the the SD of our sampling distributions for each sample size versus the ideal SEM estimate (I assume one obtained the true population SD and divided it by the square root of the sample size). The graphs are quite similar but not exactly identical. The random error in each of our trials is the cause of this discrepancy. Formally, the SEM is assumed to be in the limit as the number of repetitions approaches infinity. However, we can see that even with relatively few repetitions of the experiment we are quite close and should converge on the SEM formula within enough repetitions.

good narrative

good meaningful titles & axis labels ☺

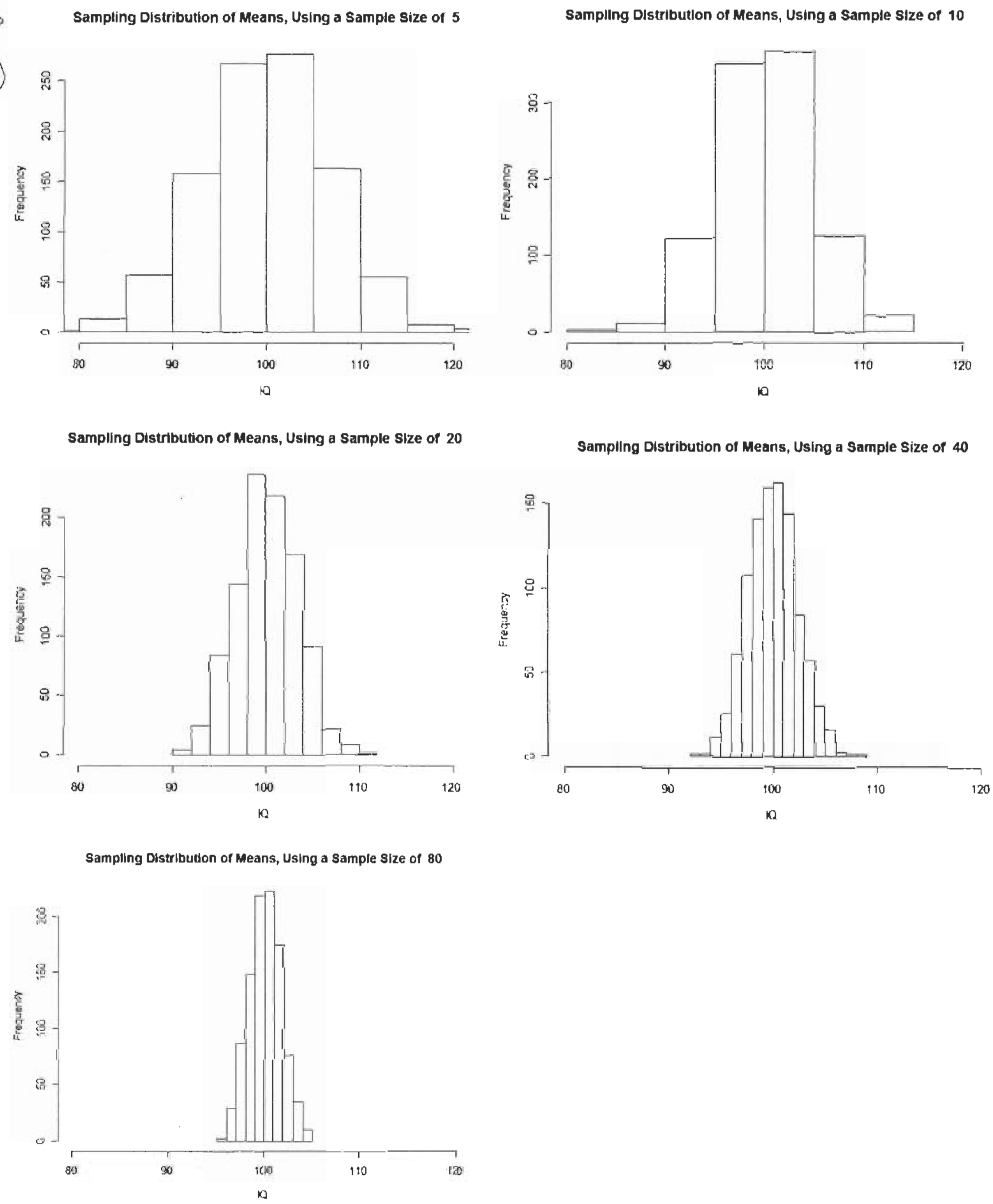
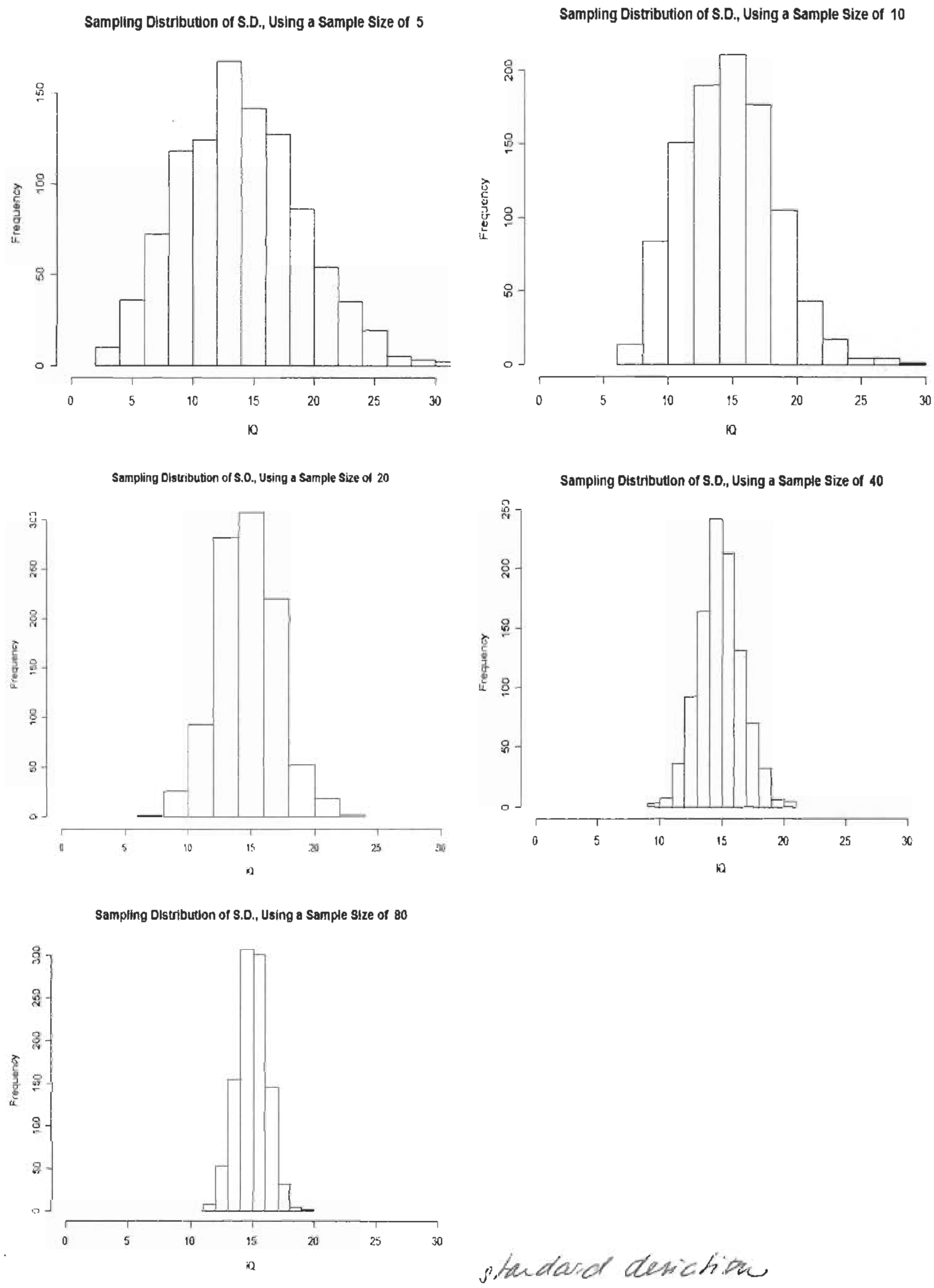
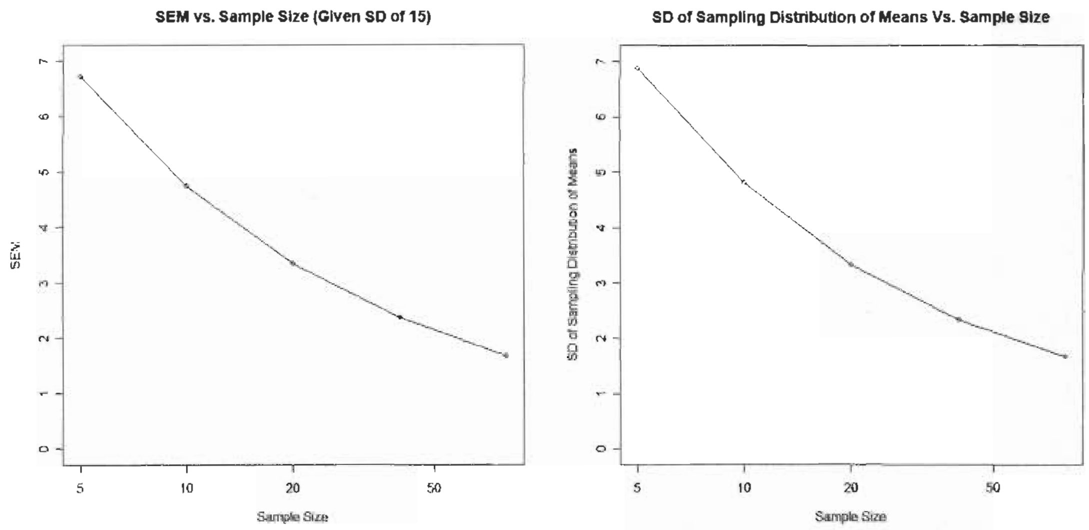


***Figure 2:** Sampling distribution of the mean, collected from sample sizes of 5,10,20,40 and 80 over 1000 repetitions*



standard deviation

Figure 3: *Sampling distribution of the ~~mean~~, collected from sample sizes of 5,10,20,40 and 80 over 1000 repetitions*



← nice descriptive titles & axis labels

would be great if you could put these in to 1 chart for easier visual comparison

Figure 4: *Comparison of the Standard error of the mean versus Standard Deviation of the sampling distributions generated with varying sample size over 1000 repetitions* ✓