STA362: Introduction to Statistics

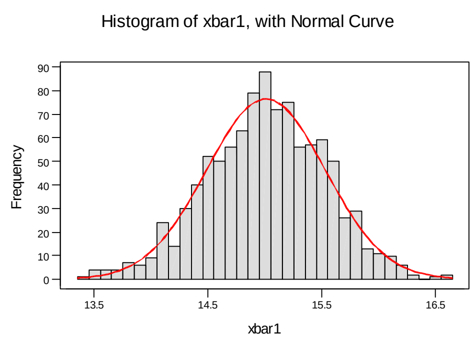
Illustration of the sampling distribution of sums and differences of sample means

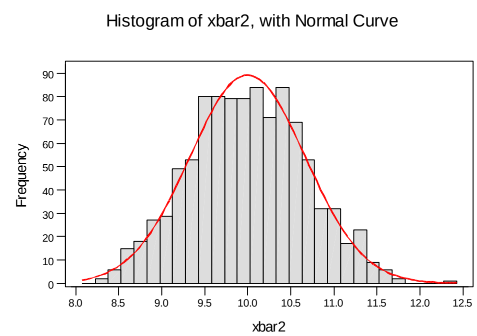
We can make inferential comparisons of two population means by looking at their difference $\mu_1 - \mu_2$. It would make intuitive sense to use $\bar{x}_1 - \bar{x}_2$ as our point estimator. The question is: what is the distribution of $\bar{x}_1 - \bar{x}_2$? Is it an unbiased estimator of $\mu_1 - \mu_2$? What is the point estimator's variance? These things are necessary to know when wishing to make a confident inference about $\mu_1 - \mu_2$.

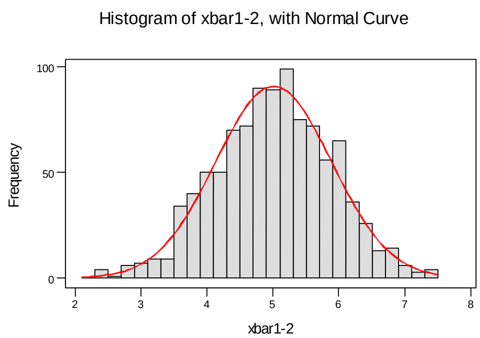
We will develop the answers to these questions in class, but it helps to see them substantiated via a computer simulation. Suppose we have two populations:

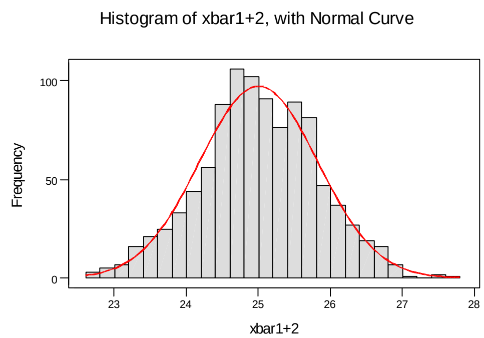
population 1: normal with $\mu_1 = 15.0$, $\sigma_1 = 3.0$
population 2: normal with $\mu_2 = 10.0$, $\sigma_2 = 4.0$

We collected 1000 simulated samples of size $n_1 = n_2 = 36$ from each population and computed four statistics in each simulation: $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_1 - \bar{x}_2$, and $\bar{x}_1 + \bar{x}_2$. Our goal here is chiefly to see how the difference and sum of two sample means behaves. Here are histograms picturing the simulated sampling distributions for each:









Descriptive Statistics for each sampling distribution

	reps	Mean	StDev	Min	Max
xbar1	1000	15.009	0.522	13.359	16.590
xbar2	1000	9.9967	0.670	8.227	12.360
xbar1-xbar2	1000	5.0119	0.878	2.331	7.462
xbar1+xbar2	1000	25.005	0.821	22.698	27.747