So far, we have only focused on estimating the mean living area in homes in Ames. Now you'll try to estimate the mean home price.  
**1. Take a random sample of size 50 from price. Using this sample, what is your best point estimate of the population mean?**

|  |
| --- |
| salePrice <- ames$SalePrice priceSample <- sample(salePrice, 50) mean(priceSample) |

**2. Since you have access to the population, simulate the sampling distribution for $\bar{x}\_{price}$ price (sampling mean of price ) by taking 5000 samples from the population of size 50 and computing 5000 sample means. Store these means in a vector called  sample\_means50 . Plot the data, then describe the shape of this sampling distribution.**

|  |
| --- |
| sample\_mean50 <- rep(NA, 5000) for(i in 1:5000){  sample50 <- sample(salePrice, 50)  sample\_mean50[i] <- mean(sample50) } hist(sample\_mean50, breaks = 20, main = "Histogram of 5000 sample[50] means", xlab = "sample mean")    Based on this sampling distribution, I can guess the mean home price is about 178000  Finally, calculate and report the population mean. |

**3.Change your sample size from 50 to 150, then compute the sampling distribution using the same method as above, and store these means in a new vector called sample\_means150 . Describe the shape of this sampling distribution, and compare it to the sampling distribution for a sample size of 50.**

|  |
| --- |
| sample\_mean150 <- rep(NA, 5000) for(i in 1:5000){  sample150 <- sample(salePrice, 150)  sample\_mean150[i] <- mean(sample150) } hist(sample\_mean150, breaks = 20, main = "Histogram of 5000 sample[150] means", xlab = "sample mean[150]")    Compare to sample 50, mean of sample 150 looks a little difference. It’s easy to see that when we pick up larger sample size, the shape becomes more precise and closer to mean(salePrice).  Base on that, I can guest the mean home price is about 179500. |

**4. Of the sampling distributions from 2 and 3, which has a smaller spread?**  
If we're concerned with making estimates that are more often close to the  
true value, would we prefer a distribution with a large or small spread?

The larger sample size (150) has smaller spread. And the estimate is closer to true value.