Data Science Module 5 Exercises

Brady Lamson

2/28/2022

9: Statistical Foundations

9.3: Simulations

Exercise 1:

```
# A
sample_mean_vec <- c()

for(i in 1:1000) {
    sim_sample <- rnorm(n = 10, mean = 50, sd = 15)
    sample_mean_vec <- c(sample_mean_vec, mean(sim_sample))
}
```

```
#B
sample_mean <- mean(sample_mean_vec) %>% round(digits = 3)
sample_standard_error <- sd(sample_mean_vec) %>% round(digits = 3)

glue::glue("
    The mean of the sample mean vector is approximately {sample_mean},
    and the standard error of the vector is approximately {sample_standard_error}.
")
```

```
## The mean of the sample mean vector is approximately 49.995,
## and the standard error of the vector is approximately 4.904.
```

c) We can see that both values we got are very close to their theoretical values. The theoretical mean would be 50 and we can calculate the theoretical standard error, $\sigma/\sqrt{n} = 15/\sqrt{10} = 4.74342$.

```
# D

ggplot(data = data.frame(sample_mean_vec)) +
    geom_histogram(
        mapping = aes(x = sample_mean_vec, y = stat(density)),
        binwidth = 1,
        color = "black") +
    geom_function(
        fun = dnorm,
```

```
        args = list(mean = 50, sd = 15/sqrt(10)),
        color = "blue") +
    labs(title = "Sampling Distribution")
```