

Matt Fertakos  
Worked with Bonnie

**Q1:**

```
x=dat_pen$body_mass_g  
n_1= 68  
n_2= 152
```

```
rm(list = ls())  
sse_mean = function(x)  
{  
  sd(x,na.rm=TRUE)/sqrt(length(x)-sum(is.na(x)))  
}  
sse_mean(penguins$body_mass_g)  
sse_mean(mtcars$mpg)
```

**Q2:**

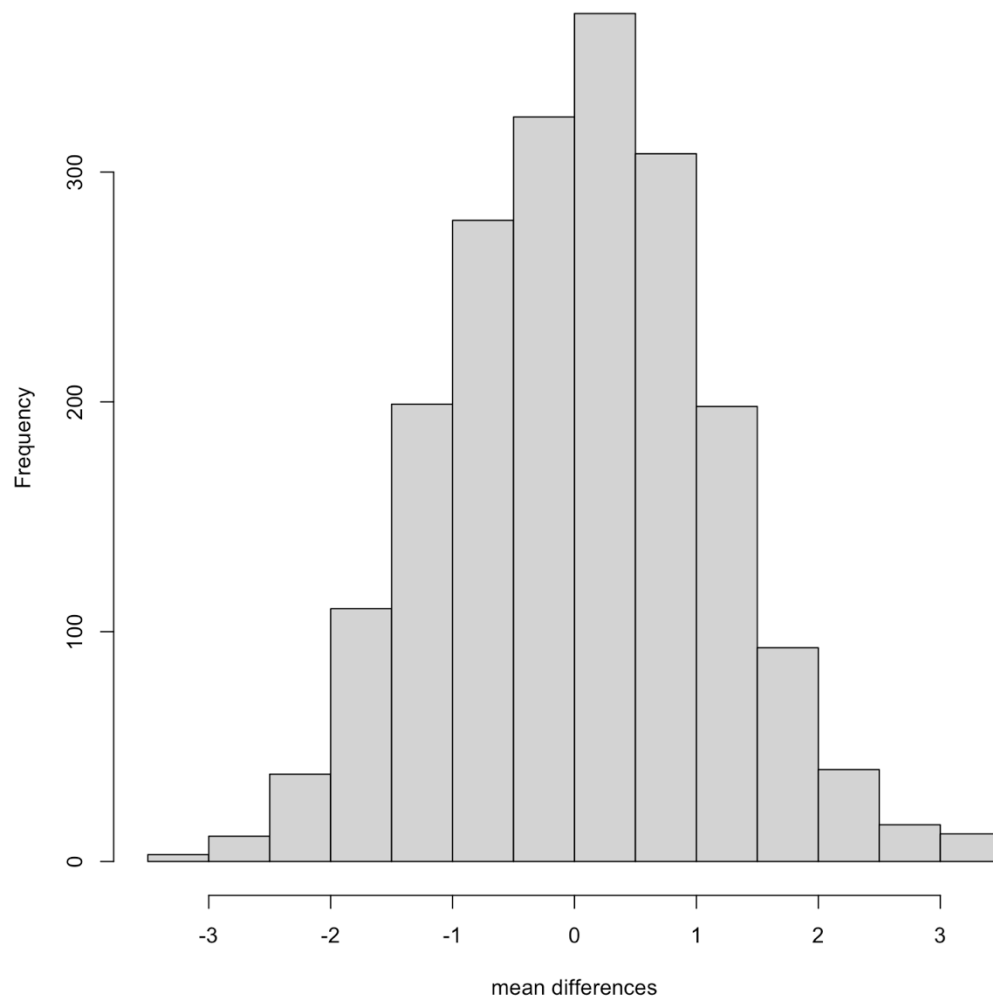
```
two_group_resample = function(x, n_1, n_2)  
{  
  dat_1 = sample(x, n_1, replace = TRUE)  
  dat_2 = sample(x, n_2, replace = TRUE)  
  difference_in_means = mean(dat_1,na.rm = TRUE) - mean(dat_2,na.rm=TRUE)  
  return(difference_in_means)  
}
```

**Q3:**

This function simulates the null hypothesis because it takes a random number of points and shuffles them to create a random dataset. This is an example of Monte Carlo sampling because in the sampling associations between the flipper length and species are destroyed. The observed data is then compared to this to see if the trend is significant.

**Q4:**

**Histogram of random resampled differences of mean flipper lengths (mm)  
between Adelie and Chinstrap penguins**



**Q5:**

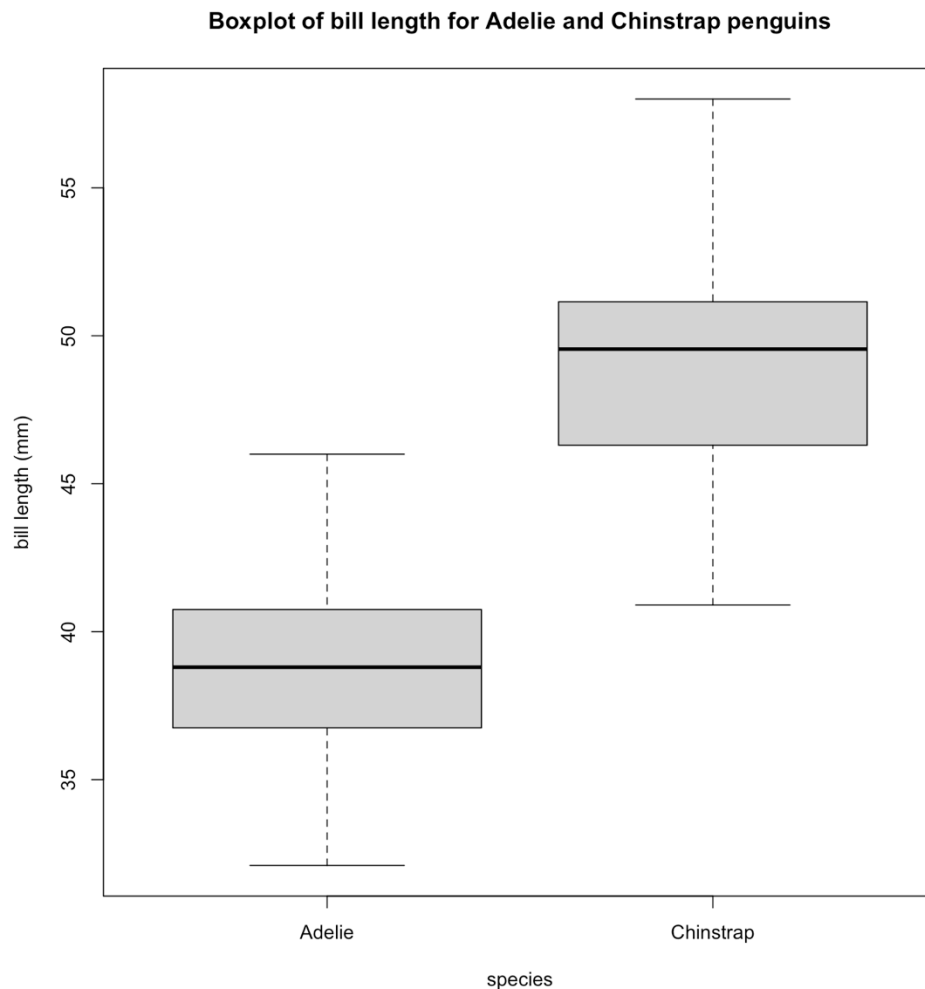
`mean_differences[which(abs(mean_differences)>5.8)]`

None are greater than 5.8 (or lower than -5.8)

**Q6:**

At least 10 million or more simulations. This is because there is such a small chance to get this result when a random sampling of flipper lengths is completed.

**Q7:**



**Q8:**

Group means: 38.79139 & 48.83382

diff\_crit = 10.04 (difference between the means)

**Q9:**

The p value is less than  $2.2 \times 10^{-16}$ . This value means that there is an EXTREMELY small chance that the result obtained by the observed values would occur if there is in fact no association between bill length and species. This chance is equal to less than 1 in a quadrillion. As a result, we can interpret this as there being a significant difference between bill length in these two species.

**Q10:**

There were 0 differences in means greater than  $\text{diff\_crit}$ , which is equal to 10.04. This means in a random sampling with two different sample sizes (68 and 152) repeated 1000 times, there was no occurrence where the difference in mean bill length between the two sample sizes was greater than or equal to the observed difference in mean bill length of 10.04. The repeating creates a distribution of possible differences in mean bill lengths when the species component is destroyed (a null distribution). 10.04 does not fall within this distribution, which was calculated between the differences in means of each species.

**Q11:**

**Histogram of random resampled differences in mean bill lengths  
of Adelie and Chinstrap penguins**

