

Midterm2 Review

Nate Wells

4/2/2021

```
penguins <- penguins %>% drop_na()
```

Do Adelie and Gentoo penguins have different body mass?

Let μ_a average body mass for Adelie penguins and μ_g the average body mass for Gentoo penguins.

$$H_0 : \mu_a = \mu_g$$

Alt Hypothesis: Choose one of the following

Choice I

$$H_0 : \mu_a \neq \mu_g$$

Choice II

$$H_0 : \mu_a > \mu_g$$

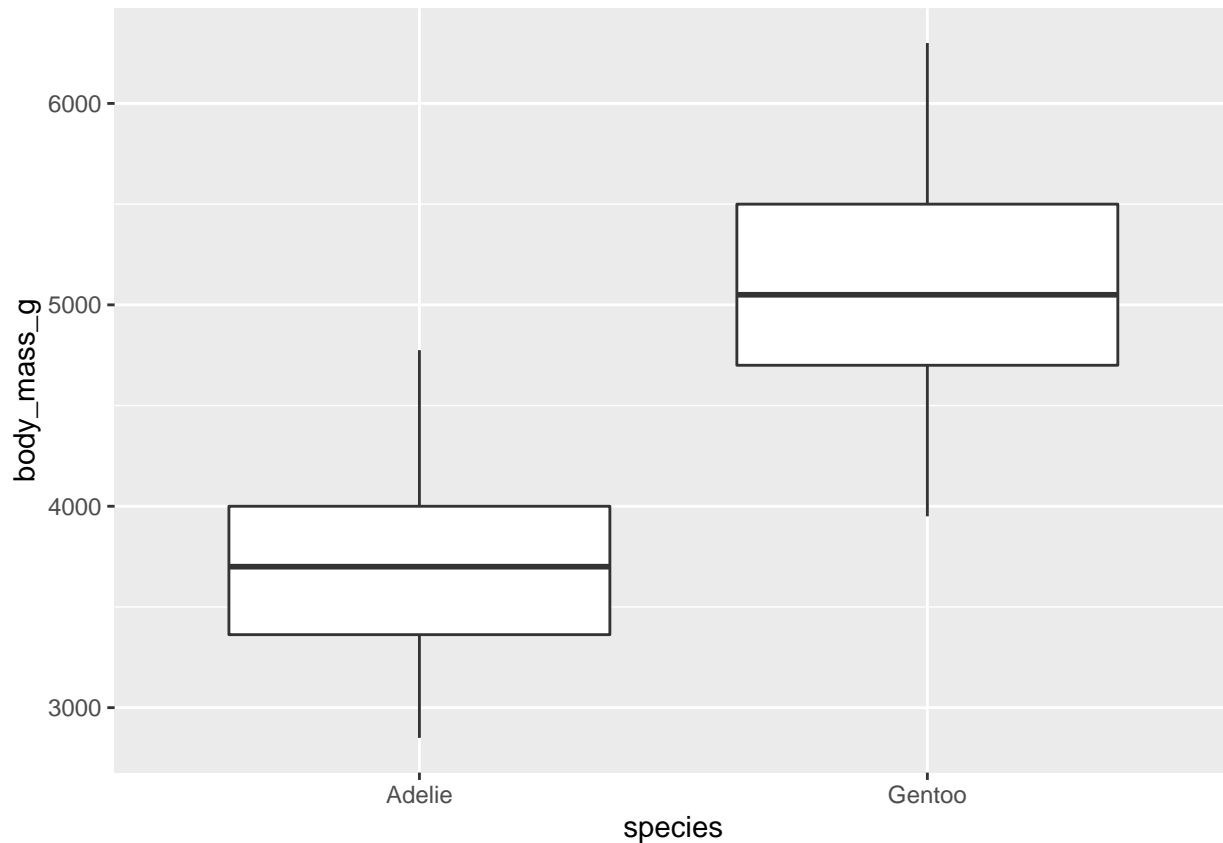
Choice III

$$H_0 : \mu_a < \mu_g$$

I choose choice 1

```
penguins_ag <- penguins %>% filter(species %in% c("Adelie", "Gentoo"))
```

```
ggplot(penguins_ag, aes(x = species, y= body_mass_g ))+geom_boxplot()
```



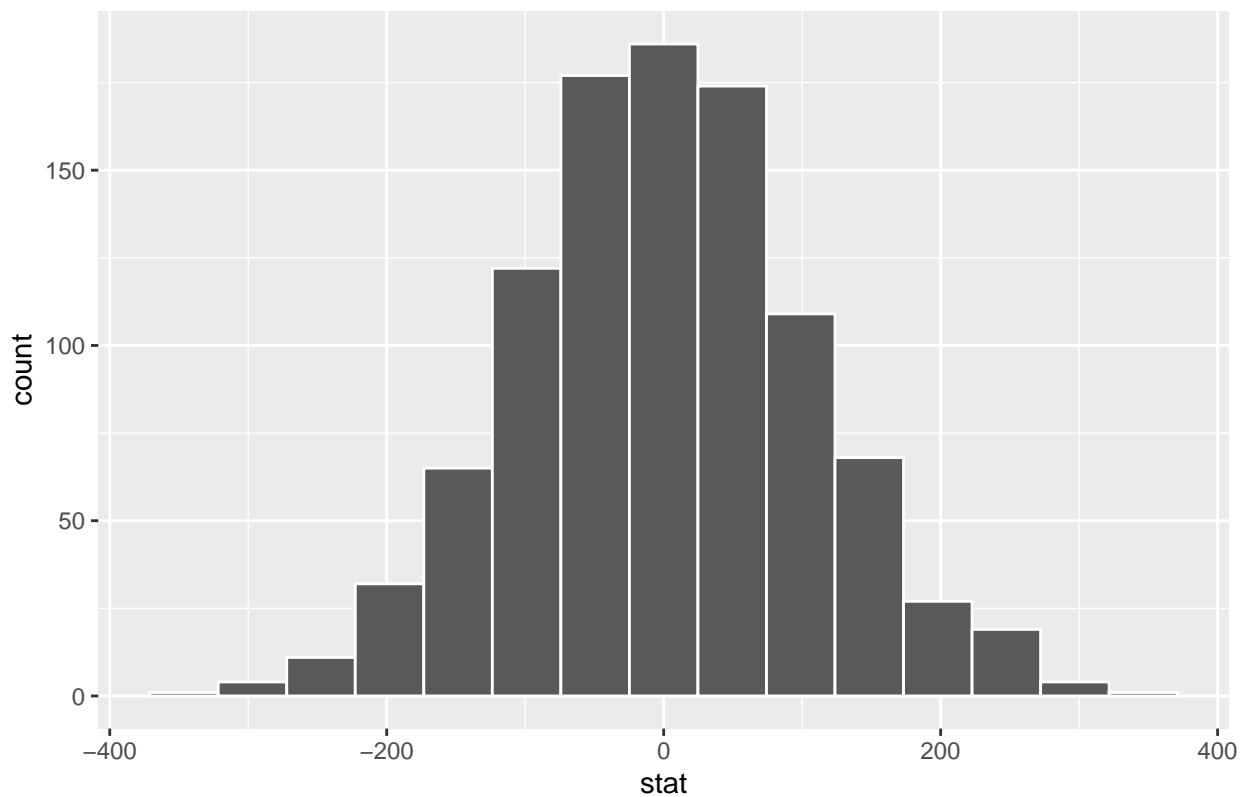
Is it likely there is a difference?

```
penguin_stats<-penguins_ag %>%  
  specify(response = body_mass_g, explanatory = species) %>%  
  hypothesise(null = "independence") %>%  
  generate(reps = 1000, type = "permute") %>%  
  calculate(stat = "diff in means", order = c("Adelie", "Gentoo"))
```

Other types: 1. If we have null = "point" and response is quantitative, use type = "bootstrap" 2. If we have null = "point" and response is categorical, use type = "simulate" 3. If doing confidence intervals, always use type = "bootstrap"

```
penguin_stats %>% visualise()
```

Simulation-Based Null Distribution



```
og_stat <- penguins_ag %>%  
  specify(response = body_mass_g, explanatory = species) %>%  
  calculate(stat = "diff in means", order = c("Adelie", "Gentoo"))
```

```
og_stat
```

```
## # A tibble: 1 x 1  
##   stat  
##   <dbl>  
## 1 -1386.
```

```
penguin_stats %>% get_p_value(obs_stat = og_stat, direction="both")
```

```
## Warning: Please be cautious in reporting a p-value of 0. This result is an  
## approximation based on the number of `reps` chosen in the `generate()` step. See  
## `?get_p_value()` for more information.
```

```
## # A tibble: 1 x 1  
##   p_value  
##   <dbl>  
## 1      0
```

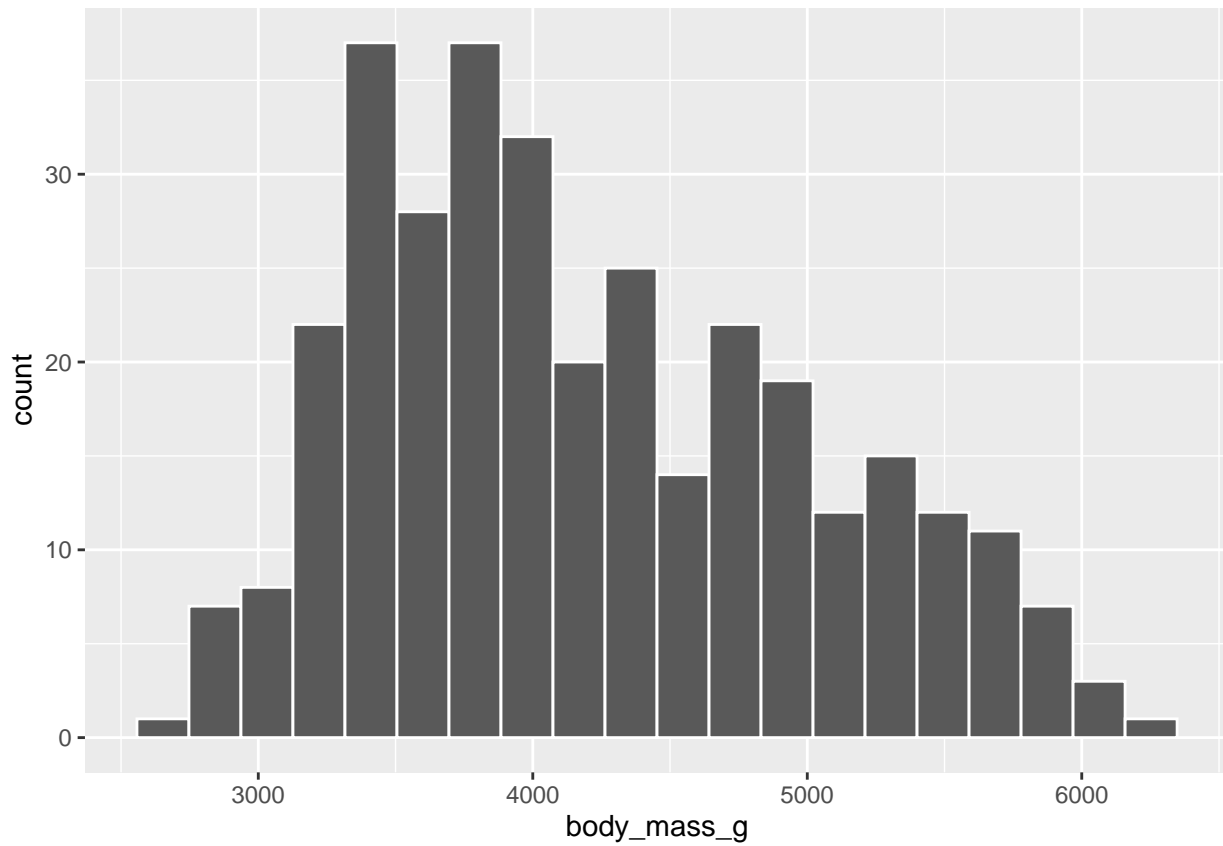
Make conclusion by comparing p-value to significance level α . Standard $\alpha = 0.05$. Since $p\text{-value} < 0.05$, we have good evidence to reject H_0 in favor of H_a .

Making a Sampling and Bootstrap Distribution

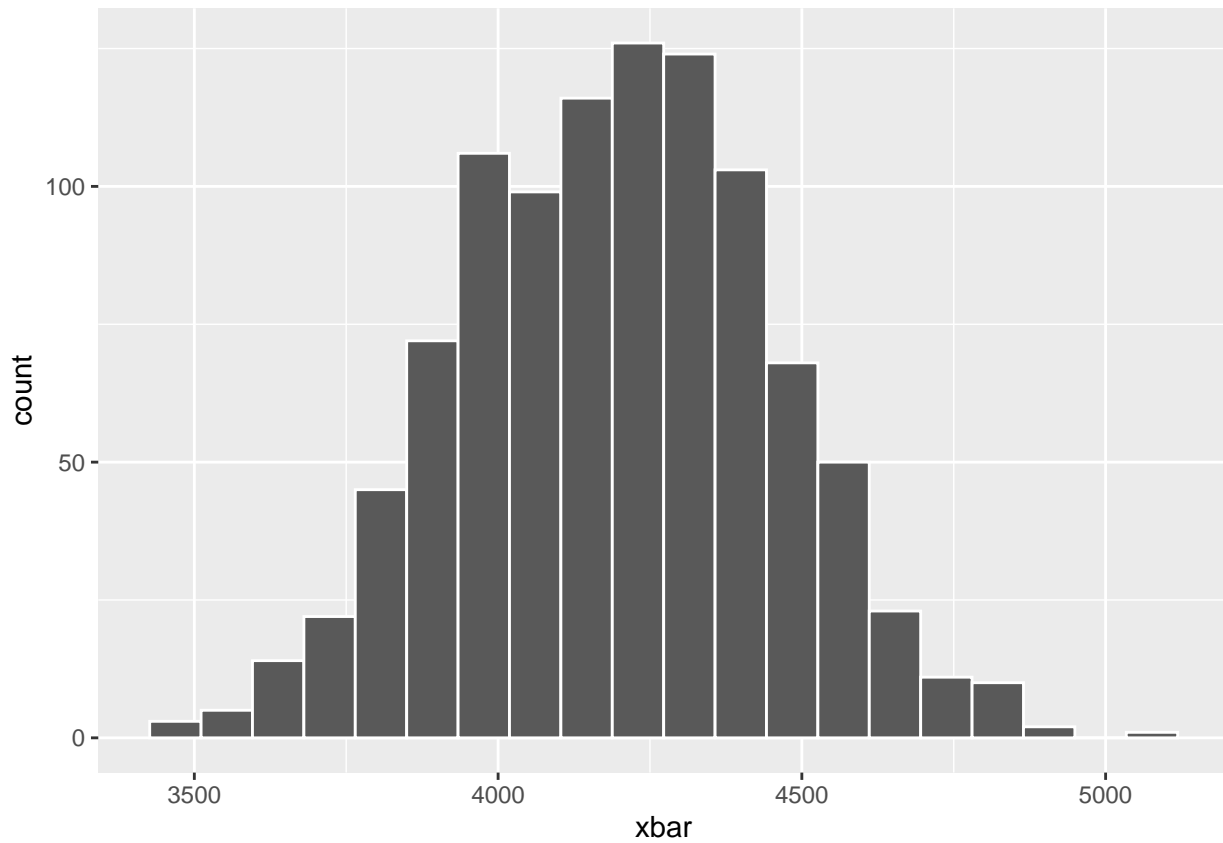
```
penguin_samp <- penguins %>%  
  rep_sample_n(size = 10, replace = F, reps = 1000) %>%
```

```
group_by(replicate) %>% summarize(xbar = mean(body_mass_g))
```

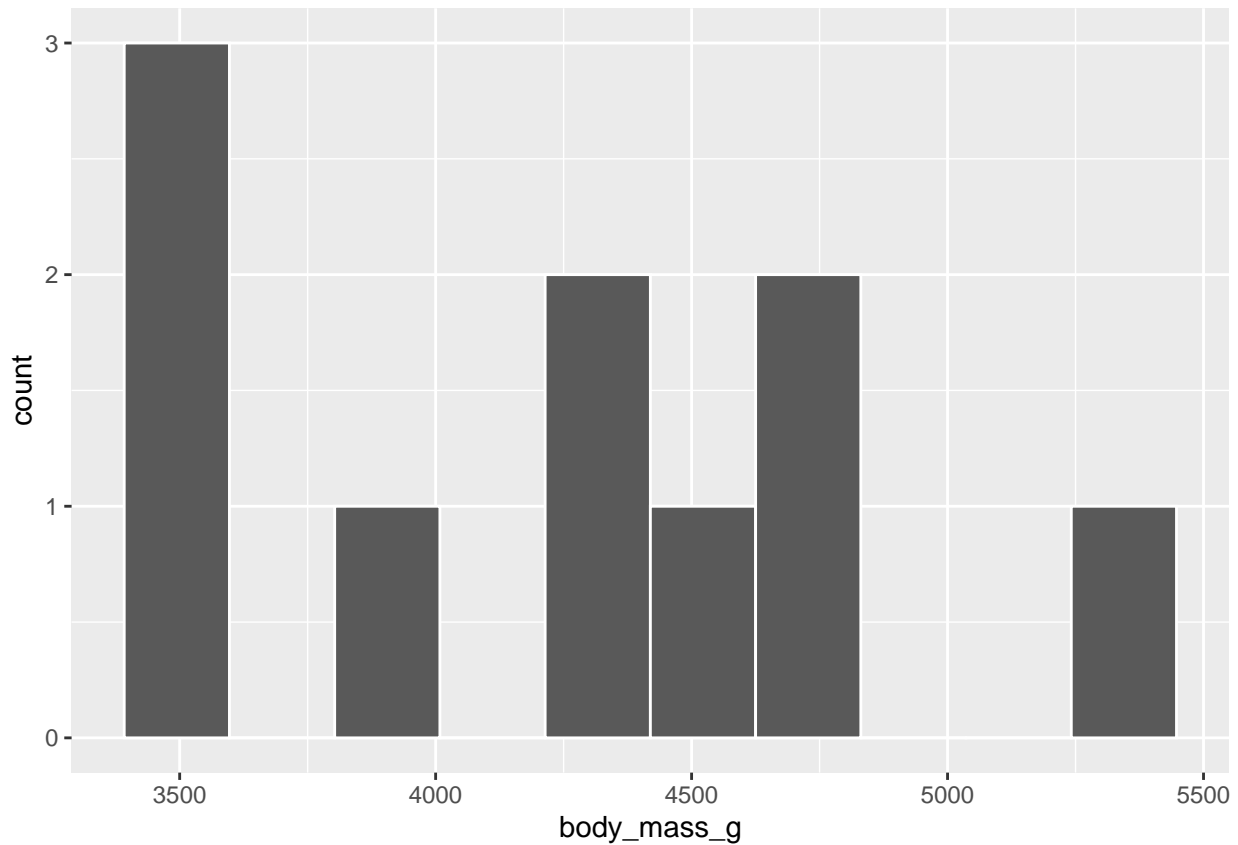
```
ggplot(penguins, aes(x = body_mass_g))+geom_histogram(bins = 20, color = "white")
```



```
ggplot(penguin_samp, aes(x = xbar))+geom_histogram(bins = 20, color = "white")
```



```
penguin_sample <- penguins %>% sample_n( size = 10, replace = F)
ggplot(penguin_sample, aes(x = body_mass_g))+geom_histogram(bins = 10, color = "white")
```



```
penguin_boot <- penguin_sample %>%  
  rep_sample_n( size = 10, replace = T, reps = 1000) %>%  
  group_by(replicate) %>% summarize(xbar = mean(body_mass_g))
```

```
ggplot(penguin_boot, aes(x = xbar))+geom_histogram(bins = 20, color = "white")
```

