

Penguin Assignment Q4

I have performed a t-test to determine whether the mean bodymass of gentoo and chinstrap penguins differ significantly from each other and have produced an overlapping histogram of this data to visually illustrate this test. My work is presented below in three overall steps: loading and cleaning the data; performing the t test; and producing and exporting the plot.

```
setwd("~/Documents/Year 3/penguin_assessment_Q4") #must be altered to individual's machine  
source("functions/libraries.R") #loading in required libraries
```

Loading and cleaning the data

```
##  
## Attaching package: 'janitor'  
## The following objects are masked from 'package:stats':  
##  
##   chisq.test, fisher.test  
##  
## Attaching package: 'dplyr'  
## The following objects are masked from 'package:plyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
source("functions/functions.R") #loading in functions  
source("functions/plotting.R") #loading in functions related to plotting  
  
write.csv(penguins_raw, "data_raw/penguins_raw.csv") #preserving a read only version of the raw data  
  
penguin_data_clean <- data_cleaning(penguins_raw) #cleaning data  
write.csv(penguin_data_clean, "data_clean/penguin_data_clean.csv") #preserving copy of clean data  
  
removed_empty_bodymass <- remove_empty_bodymass(penguin_data_clean) #remove N/A bodymass values  
  
gentoo_and_chinstrap <- remove_adelie(removed_empty_bodymass) #removing adelie species from dataset
```

Performing the T-test In order to determine what type of T-test should be used one must first establish whether the data has a normal distribution and whether the two data sets have equal variance or not.

Shapiro-Wilk test for normality of gentoo data:

H0: data is normal H1: data is not normal Significance level = 0.05

```
#shapiro-wilk test to test normality of gentoo bodymass data
```

```
shapiro_wilk_gentoo <- with(gentoo_and_chinstrap, shapiro.test(body_mass_g[species == "Gentoo penguin (Pygoscelis papua)"]))
shapiro_wilk_gentoo
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: body_mass_g[species == "Gentoo penguin (Pygoscelis papua)"]
```

```
## W = 0.98593, p-value = 0.2336
```

p = 0.2336, 0.2336 > 0.05 therefore accept H0. Data is normal

Shapiro-Wilk test for normality of chinstrap data:

H0: data is normal H1: data is not normal Significance level = 0.05

```
shapiro_wilk_chinstrap <- with(gentoo_and_chinstrap, shapiro.test(body_mass_g[species == "Chinstrap penguin (Pygoscelis antarctica)"]))
shapiro_wilk_chinstrap
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: body_mass_g[species == "Chinstrap penguin (Pygoscelis antarctica)"]
```

```
## W = 0.98449, p-value = 0.5605
```

p = 0.5605, 0.5605 > 0.05 therefore accept H0 Data is normal

Test whether variance is unequal:

H0: variance is equal H1: variance is unequal Significance level = 0.05

```
var.test(body_mass_g ~ species, data = gentoo_and_chinstrap)
```

```
##
```

```
## F test to compare two variances
```

```
##
```

```
## data: body_mass_g by species
```

```
## F = 0.58124, num df = 67, denom df = 122, p-value = 0.01559
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.3852832 0.9004223
```

```
## sample estimates:
```

```
## ratio of variances
```

```
## 0.5812443
```

p = 0.01559, 0.01559 < 0.05 therefore reject H0. Variance is not equal

Therefore we cannot use student T test. Welch's t test will be used:

H0: mean bodymass of gentoo and chinstrap penguins are equal H1: mean bodymass of gentoo and chinstrap penguins are not equal Significance level = 0.05

```
welch_t_test <- t.test(body_mass_g ~ species, data = gentoo_and_chinstrap, var.equal = FALSE)
welch_t_test
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: body_mass_g by species
## t = -20.628, df = 170.4, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Chinstrap penguin (Pygoscelis antarctica) and group Gentoo penguin (Pygoscelis papua) is not equal to 0
## 95 percent confidence interval:
## -1471.440 -1214.416
## sample estimates:
## mean in group Chinstrap penguin (Pygoscelis antarctica)
## 3733.088
## mean in group Gentoo penguin (Pygoscelis papua)
## 5076.016
```

$p = 2.2e-16$, $2.2e-16 < 0.05$ therefore reject H_0 . Mean bodymass of gentoo and chinstrap penguins are not equal.

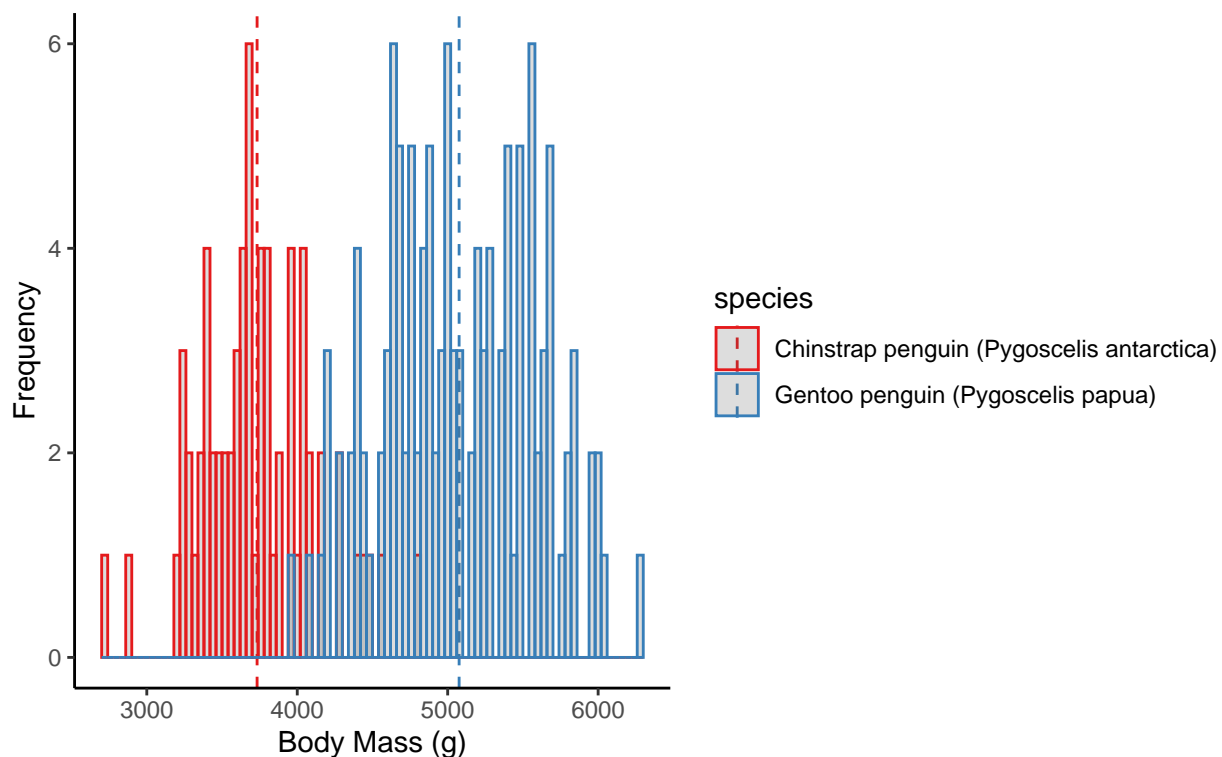
Producing and exporting the plot

```
plot_bodymass_figure(gentoo_and_chinstrap) #plotting figure
```

```
## Warning: Use of `gentoo_and_chinstrap$body_mass_g` is discouraged.
## i Use `body_mass_g` instead.
## Warning: Use of `gentoo_and_chinstrap$species` is discouraged.
## i Use `species` instead.
```

Gentoo and Chinstrap Penguin bodymass distributions

T-test visual



```
save_bodymass_plot_png(gentoo_and_chinstrap) #exporting figure as pdf
```

```
## Warning: Use of `gentoo_and_chinstrap$body_mass_g` is discouraged.
## i Use `body_mass_g` instead.
## Use of `gentoo_and_chinstrap$species` is discouraged.
```

```
## i Use `species` instead.
```

```
## pdf
```

```
## 2
```

```
save_bodymass_plot_svg(gentoo_and_chinstrap) #exporting figure as svg
```

```
## Warning: Use of `gentoo_and_chinstrap$body_mass_g` is discouraged.
```

```
## i Use `body_mass_g` instead.
```

```
## Use of `gentoo_and_chinstrap$species` is discouraged.
```

```
## i Use `species` instead.
```

```
## pdf
```

```
## 2
```

NB: Please view functions.R and plotting.R files for code used to produce cleaning and plotting functions.