

Using_Models_2

Ethan Rutledge

11/27/2021

1-sample t-test on the Gentoo penguin flipper lengths = 0

```
require(palmerpenguins)
```

```
## Loading required package: palmerpenguins
```

```
t.test(subset(penguins, species == "Gentoo")$flipper_length_mm)
```

```
##  
## One Sample t-test  
##  
## data: subset(penguins, species == "Gentoo")$flipper_length_mm  
## t = 371.43, df = 122, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 216.0295 218.3445  
## sample estimates:  
## mean of x  
## 217.187
```

1-sample t-test on the Gentoo penguin flipper lengths = 218

```
t.test(  
  x = subset(penguins, species == "Gentoo")$flipper_length_mm,  
  mu = 218  
)
```

```
##  
## One Sample t-test  
##  
## data: subset(penguins, species == "Gentoo")$flipper_length_mm  
## t = -1.3904, df = 122, p-value = 0.1669  
## alternative hypothesis: true mean is not equal to 218  
## 95 percent confidence interval:  
## 216.0295 218.3445  
## sample estimates:  
## mean of x  
## 217.187
```

one-tailed alternative hypothesis: Gentoo penguin flippers are smaller than 218 mm

```
t.test(
  x = subset(penguins, species == "Gentoo")$flipper_length_mm,
  mu = 218,
  alternative = "less"
)
```

```
##
## One Sample t-test
##
## data: subset(penguins, species == "Gentoo")$flipper_length_mm
## t = -1.3904, df = 122, p-value = 0.08347
## alternative hypothesis: true mean is less than 218
## 95 percent confidence interval:
##      -Inf 218.1561
## sample estimates:
## mean of x
##      217.187
```

2 sample Ttest compare the flipper lengths of two species

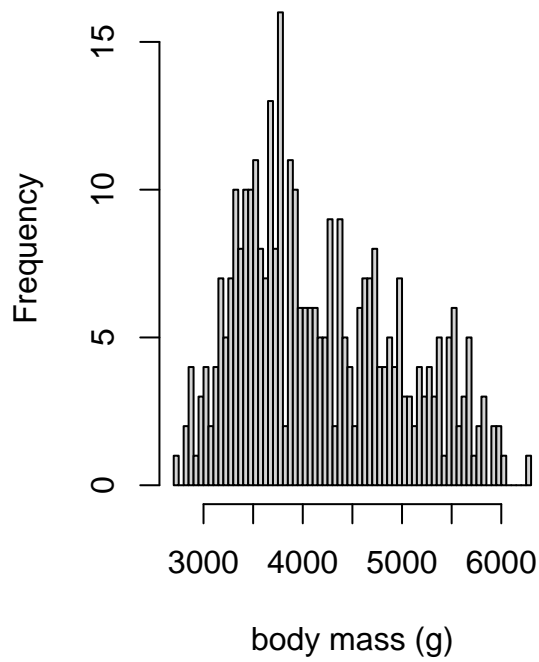
```
t.test(flipper_length_mm ~ species, data = subset(penguins, species != "Chinstrap"))
```

```
##
## Welch Two Sample t-test
##
## data: flipper_length_mm by species
## t = -34.445, df = 261.75, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Adelie and group Gentoo is not equal to 0
## 95 percent confidence interval:
##      -28.79018 -25.67652
## sample estimates:
## mean in group Adelie mean in group Gentoo
##           189.9536           217.1870
```

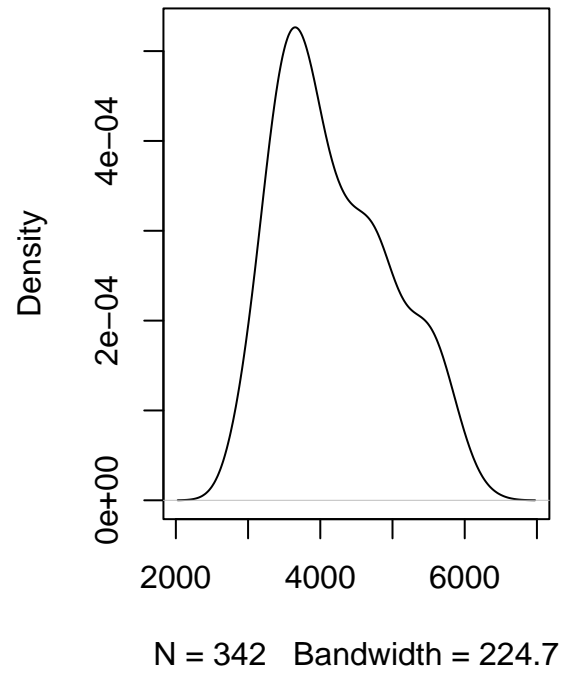
Look at the data

```
par(mfrow = c(1, 2))
hist(penguins$body_mass_g, breaks = 80, main = "histogram of body mass", xlab = "body mass (g)")
plot(density(penguins$body_mass_g, na.rm = TRUE), main = "density plot of body mass")
```

histogram of body mass

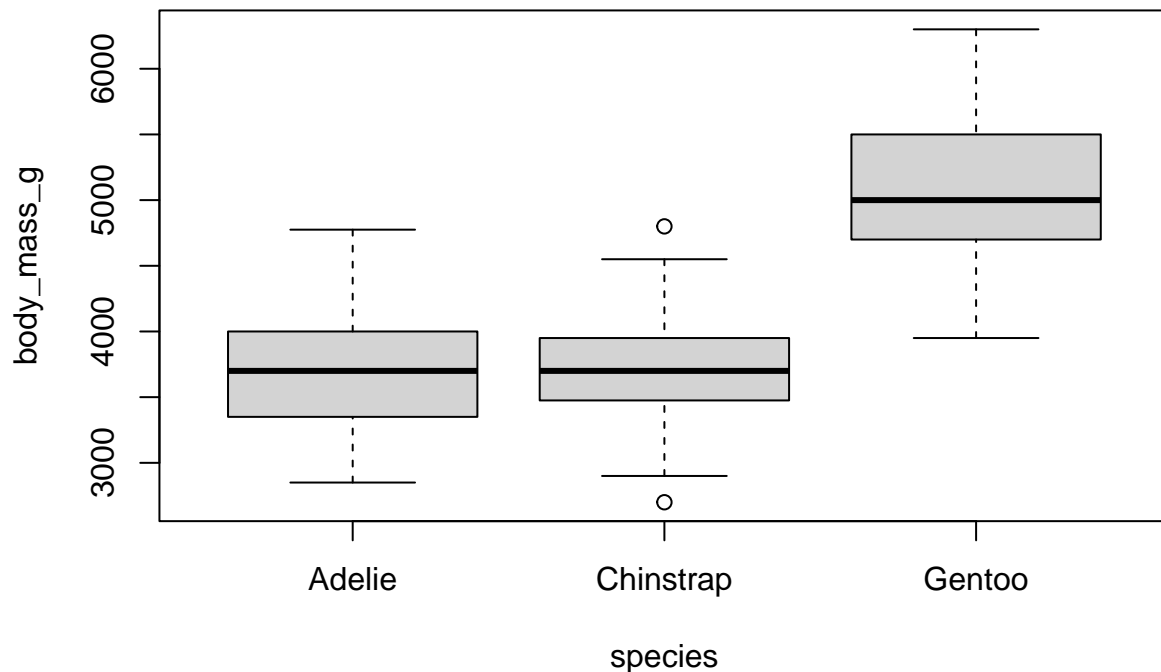


density plot of body mass



Look at the data

```
boxplot(body_mass_g ~ species, data = penguins)
```



look at the means

```
dat_chinstrap = subset(penguins, species == "Chinstrap")
mean(dat_chinstrap$body_mass_g, na.rm = TRUE)
```

```
## [1] 3733.088
```

test whether within-group data are normally-distributed:

```
shapiro.test(dat_chinstrap$body_mass_g)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dat_chinstrap$body_mass_g
## W = 0.98449, p-value = 0.5605
```

Here's a cool shortcut for calculating the species mean body masses using `aggregate()` and the formula notation

```
aggregate(body_mass_g ~ species, data = penguins, FUN = mean)
```

```
##      species body_mass_g
## 1   Adelie    3700.662
## 2 Chinstrap    3733.088
## 3   Gentoo    5076.016
```

Same thing with the shapiro test

```
aggregate(
  body_mass_g ~ species,
  data = penguins,
  FUN = function(x) shapiro.test(x)$p.value)
```

```
##      species body_mass_g
## 1   Adelie  0.03239702
## 2 Chinstrap 0.56050824
## 3   Gentoo  0.23361649
```

Fit a linear model, Then we can look at the model coefficients:

```
fit_species = lm(body_mass_g ~ species, data = penguins)
summary(fit_species)
```

```
##
## Call:
## lm(formula = body_mass_g ~ species, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1126.02  -333.09   -33.09   316.91  1223.98
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3700.66     37.62   98.37  <2e-16 ***
## speciesChinstrap    32.43     67.51    0.48   0.631
## speciesGentoo    1375.35     56.15   24.50  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 462.3 on 339 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.6697, Adjusted R-squared:  0.6677
## F-statistic: 343.6 on 2 and 339 DF, p-value: < 2.2e-16
```

ANOVA

```
anova(fit_species)
```

```
## Analysis of Variance Table
##
## Response: body_mass_g
##      Df Sum Sq Mean Sq F value    Pr(>F)
## species    2 146864214 73432107  343.63 < 2.2e-16 ***
## Residuals 339  72443483   213698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The syntax to build the model in R is easy. Use the `lm()` function to store the model in a variable: ANOVA complete walkthrough

```
fit_species = lm(body_mass_g ~ species, data = penguins)

summary(fit_species)
```

```
##
## Call:
## lm(formula = body_mass_g ~ species, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1126.02  -333.09   -33.09   316.91  1223.98
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3700.66      37.62   98.37  <2e-16 ***
## speciesChinstrap    32.43      67.51    0.48   0.631
## speciesGentoo    1375.35      56.15   24.50  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 462.3 on 339 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.6697, Adjusted R-squared:  0.6677
## F-statistic: 343.6 on 2 and 339 DF, p-value: < 2.2e-16
```

```
anova(fit_species)
```

```
## Analysis of Variance Table
##
## Response: body_mass_g
##           Df Sum Sq Mean Sq F value    Pr(>F)
## species     2 146864214 73432107  343.63 < 2.2e-16 ***
## Residuals  339  72443483   213698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2 Way ANOVA, without interaction terms

```
fit_additive = lm(body_mass_g ~ sex + species, data = penguins)
```

2 Way Anova, with interaction terms

```
fit_interactive = lm(body_mass_g ~ sex * species, data = penguins)

summary(fit_interactive)
```

```
##
## Call:
```

```
## lm(formula = body_mass_g ~ sex * species, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -827.21 -213.97   11.03  206.51  861.03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3368.84      36.21  93.030 < 2e-16 ***
## sexmale           674.66      51.21  13.174 < 2e-16 ***
## speciesChinstrap  158.37      64.24   2.465 0.01420 *
## speciesGentoo     1310.91      54.42  24.088 < 2e-16 ***
## sexmale:speciesChinstrap -262.89      90.85  -2.894 0.00406 **
## sexmale:speciesGentoo   130.44      76.44   1.706 0.08886 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 309.4 on 327 degrees of freedom
## (11 observations deleted due to missingness)
## Multiple R-squared:  0.8546, Adjusted R-squared:  0.8524
## F-statistic: 384.3 on 5 and 327 DF, p-value: < 2.2e-16
```

```
anova(fit_interactive)
```

```
## Analysis of Variance Table
##
## Response: body_mass_g
##           Df    Sum Sq Mean Sq F value    Pr(>F)
## sex         1 38878897 38878897 406.145 < 2.2e-16 ***
## species     2 143401584 71700792 749.016 < 2.2e-16 ***
## sex:species  2  1676557   838278   8.757 0.0001973 ***
## Residuals  327 31302628   95727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Simple Linear Regression

```
lm(bill_length_mm ~ body_mass_g, data = penguins)
```

```
##
## Call:
## lm(formula = bill_length_mm ~ body_mass_g, data = penguins)
##
## Coefficients:
## (Intercept)  body_mass_g
##  26.898872    0.004051
```