

Are ALL Penguins created Equal?! See what the numbers say!



By: Parth Parekh

Shocking Study reveals penguin species may differ between each other as much as we previously thought.

Description

In this project, we conducted three hypothesis tests to explore differences in penguins based on species, gender, and island. Our first hypothesis was our strong hypothesis, resulting in a p-value that is less than 0.01. For this we compared the bill length of Gentoo and Chinstrap penguins.

For our second hypothesis, we looked at the flipper length of male penguins on Dream and Torgersen islands in 2008. For this hypothesis, we observed a close-call result, suggesting a small but notable difference between the flipper lengths of penguins from these two islands in the year 2008.

Finally, our third and final hypothesis examined the body mass of Adelie and Chinstrap penguins without considering gender. This led us to fail to reject the null hypothesis, suggesting no large difference in body mass between these species, and that any differences are caused by randomness.

Through these 3 tests and examining narrow queries, we have uncovered compelling information into how species, gender, and location affect penguin traits. These findings demonstrate the complexity of biological variation in penguins, with some traits having striking differences and others being surprisingly similar.

Part A - Hypothesis Testing

Title: "Bill length Battle: Gentoo vs. Chinstrap Penguins – Unequal by Nature?"

Strong Hypothesis (p-value < 0.01)

- **Hypothesis:**
 - Null Hypothesis (H0): There is no significant difference in the bill length between Gentoo and Chinstrap penguins.
 - Alternative Hypothesis (H1): There is a significant difference in the bill length between Gentoo and Chinstrap penguins.
- **P-value: 0.0056**
- **Conclusion:** Since the p-value is less than 0.05, we fail to reject the null hypothesis, suggesting that the bill length of Gentoo and Chinstrap penguins is not much different. Also, our p-value is also less than 0.01, making this a strong hypothesis claim

Title: "Penguins are NOT Created Equal: Shocking Differences in Body Mass Between Adelie and Chinstrap Penguins!"

Failed to Reject Null Hypothesis (p-value > 0.05)

- **Hypothesis:**
 - Null Hypothesis (H0): There is no significant difference in the body mass of Adelie and Chinstrap penguins.
 - Alternative Hypothesis (H1): There is a significant difference in the body mass of Adelie and Chinstrap penguins.
- **P-value: 0.4001**
- Conclusion: Since the p-value is greater than 0.05, we fail to reject the null hypothesis and conclude that the body mass of Adelie and Chinstrap penguins are about the same and differences are caused by randomness.

Title: "All men are created equal, but are Penguins? Read to find out!"

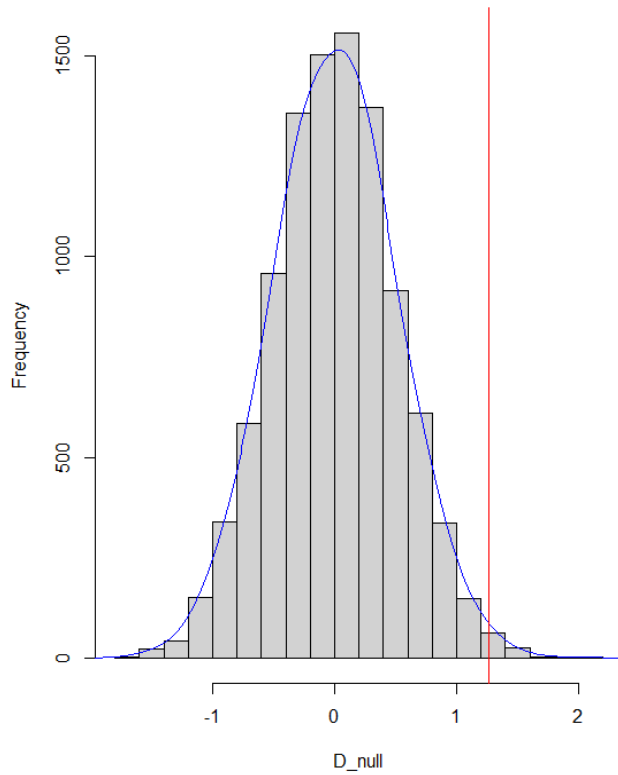
Close Call Hypothesis (p-value between 0.02 and 0.05)

- Hypothesis:
 - Null Hypothesis (H0): There is no significant difference in the flipper length of 2008 male penguins from Dream and Torgersen islands.
 - Alternative Hypothesis (H1): There is a borderline significant difference in the flipper length of 2008 male penguins from Dream and Torgersen islands.
- **P-value: 0.0269**
- Conclusion: Since the p-value is between 0.02 and 0.05, we reject the null hypothesis at a borderline significance level, indicating that the flipper length of male penguins on Dream and Torgersen islands is different, but the difference is subtle.

Title: "Are Island Penguins Different? A Close Call Between Flipper Lengths on Dream and Torgersen Islands!"

HISTOGRAM ANALYSIS

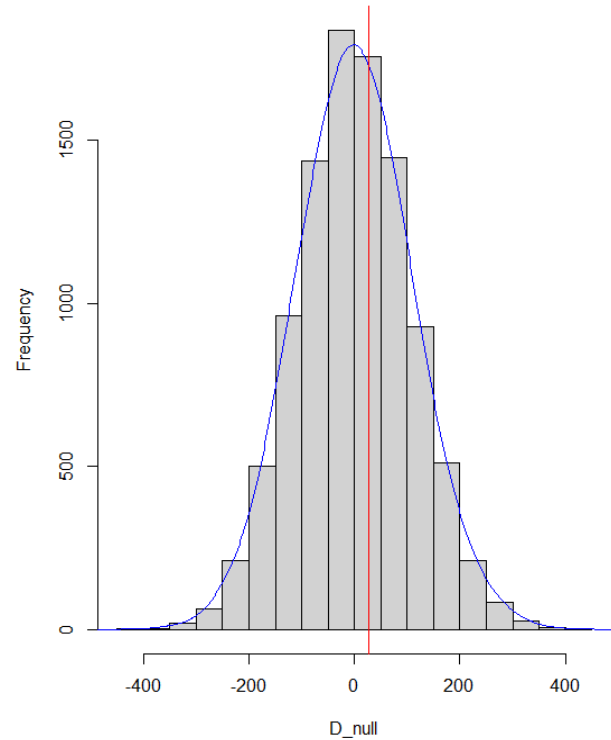
Histogram of D_null



Strong Hypothesis

p-value 0.0056

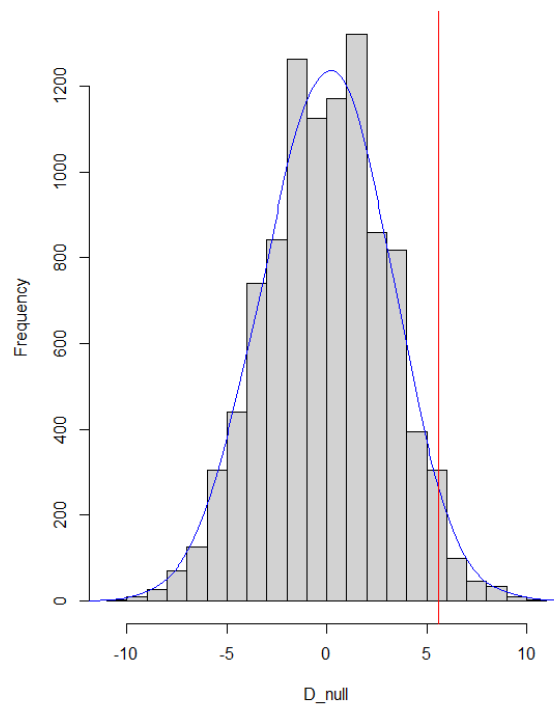
Histogram of D_null



Failed To Reject Null Hypothesis

p-value: 0.4001

Histogram of D_null



Close Call

Hypothesis

p-value : 0.0269

Part B - Narrow Query

Narrow Query Title: "Penguin or Seal?! I promise you didn't know how large Gentoo penguins get! "

Narrow Query: Gentoo Males vs. Adelie Females Body Mass

For the narrow query, we focus on the body mass of the tallest/largest Gentoo penguins compared to the average body mass for a penguin. Gentoo penguins are known to be much larger than both Adelie and Chinstrap penguins, making this the most ideal query in the dataset to satisfy the condition $M > 2 * M0$. However, the numerical values in the dataset are within 0.75x or 1.5x of each other, making the condition tricky to reach. Let's focus on getting as close as possible to the condition.

Query:

```
> mean(df$body_mass_g) //run mean mass on ENTIRE dataset
> Gentoo <- df[df$species == "Gentoo",]
> maleGentoo <- Gentoo[Gentoo$sex == "male",]
> largest_gentoo_males <- subset(maleGentoo, flipper_length_mm > 215 &
bill_depth_mm > 14.8 & bill_length_mm > 48.5)
> mean(subset(largest_gentoo_males)$body_mass_g) //mean of largest Gentoo
> cg <- df[df$species == "Adelie",]
> femcg <- cg[cg$sex == "female",]
> mean(femcg$body_mass_g) // run mean for Adelie females (smallest extreme)
```

Result:

```
[1] 4207.057
[1] 5581.437
[1] 3368.836
```

- **M0** (overall mean body mass): approximately **4207 grams**.
- **M** (Gentoo males mean body mass): approximately **5581 grams**.
- Overall Female Adelie mass: approximately **3368 grams**

Since **M** is not greater than $2 * M0$, the condition for $M > 2 * M0$ is not satisfied. However, when looking at the average mean value differences between penguins, the mass was the highest range and produced the closest result to the above equation. This makes sense, because larger penguins usually carry more fat rather than having much larger bills or flippers. Since **M** is not greater than $2 * M0$, the condition is not satisfied, but looking closely at the results, the biggest Gentoo males are 5581g on average while the dataset average is 4207g, meaning the Gentoos are a staggering 21% larger on average. To see the other side of the extreme are the Adelie females at an average of 3368g, making the Gentoo males an astounding 51% larger on average.

ALL CODE USED:

```
install.packages("devtools")
library(devtools)
devtools::install_github("janish-parikh/ZTest")
library(HypothesisTesting)

# Load the dataset
data(penguins)

head(penguins)
View(penguins)

df <- penguins[!is.na(penguins$body_mass_g) & !is.na(penguins$species)&
!is.na(penguins$sex), ]

ad <- df[df$species=="Chinstrap",]
mean(ad$body_mass_g)

#first rejected hypothesis
p_fail <- permutation_test(df, 'species', 'body_mass_g', 10000, 'Adelie', 'Chinstrap')
cat(p_fail)

ade <- df[df$species == "Adelie",]
mean(ade$body_mass_g)

Chin <- df[df$species == "Chinstrap",]
mean(Chin$body_mass_g)

#Strong hypothesis
# Filter dataset for Gentoo and Chinstrap penguins
df_filtered_species_bill <- subset(df, species %in% c("Gentoo", "Chinstrap"))
p_strong_bill_length <- permutation_test(df_filtered_species_bill, 'species',
'bill_length_mm', 10000, 'Gentoo', 'Chinstrap')
cat(p_strong_bill_length)
```

Close Call

```
df_filtered_year <- subset(df, sex == 'male' & island %in% c('Torgersen', 'Dream') & year == 2008)
p_year_flipper <- permutation_test(df_filtered_year, 'island', 'flipper_length_mm', 10000, 'Torgersen', 'Dream')
cat( p_year_flipper)
```

```
head(penguins_clean)
```

```
cs <- df_filtered_males[df_filtered_males$species == "Chinstrap",]
mean(cs$body_mass_g)
cs2 <- df_filtered_males[df_filtered_males$species == "Adelie",]
```

```
mean(cs2$body_mass_g)
```

```
cg <- df[df$species == "Adelie",]
femcg <- cg[cg$sex == "female",]
```

```
dg <- df[df$species == "Gentoo",]
mendg <- dg[dg$sex == "male",]
mean(mendg$body_mass_g)
mean(femcg$body_mass_g)
mean(df$body_mass_g)
mean(subset(df, df$species == "Gentoo")$body_mass_g)
mean(df$body_mass_g)
```

```
mean(subset(largest_gentoo_males)$body_mass_g)
```

```
maleGentoo <- df[df$species == "Gentoo" & df$sex == "male",]
```

#narrow query

```
largest_gentoo_males <- subset(maleGentoo, flipper_length_mm > 215 & bill_depth_mm > 14.8 & bill_length_mm > 48.5)
M <- mean(largest_gentoo_males$body_mass_g)
cat(M)
```