Exam 02 Review

2019-11-21

```
library(Sleuth3)
library(knitr)
library(broom)
library(tidyverse)
library(nnet)
library(cowplot)
```

Data Description

- We would like to identify crab species based on the closing force and propodus height of claws
  - `ex0722` data set in the `Sleuth3` R package
- **Predictors:**
  - `Force` : Closing force of claw (newtons)
  - `Height` : Propodus height (mm)
- **Response:**
  - `Species` : Hemigrapsus nudus (Hn), Lophopanopeus bellus (Lb), Cancer productus (Cp)

Part I: Lb species?

Suppose we want to use `Force` and `Height` to determine whether or not a crab is from the Lophopanopeus bellus (Lb) species.

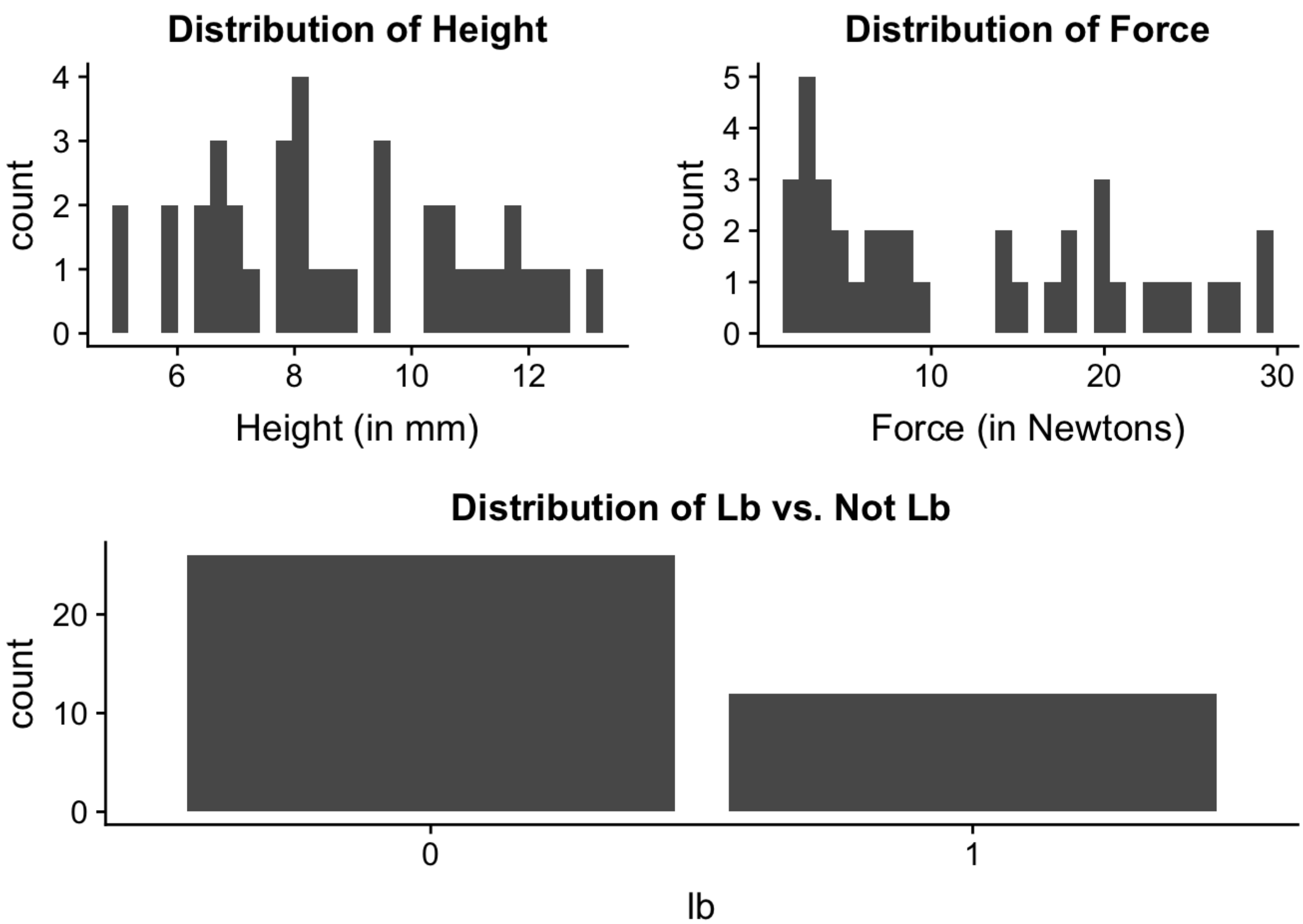
1. What type of model should we use? Briefly explain your choice.

Logistic regression, because the response is categorical with two levels.

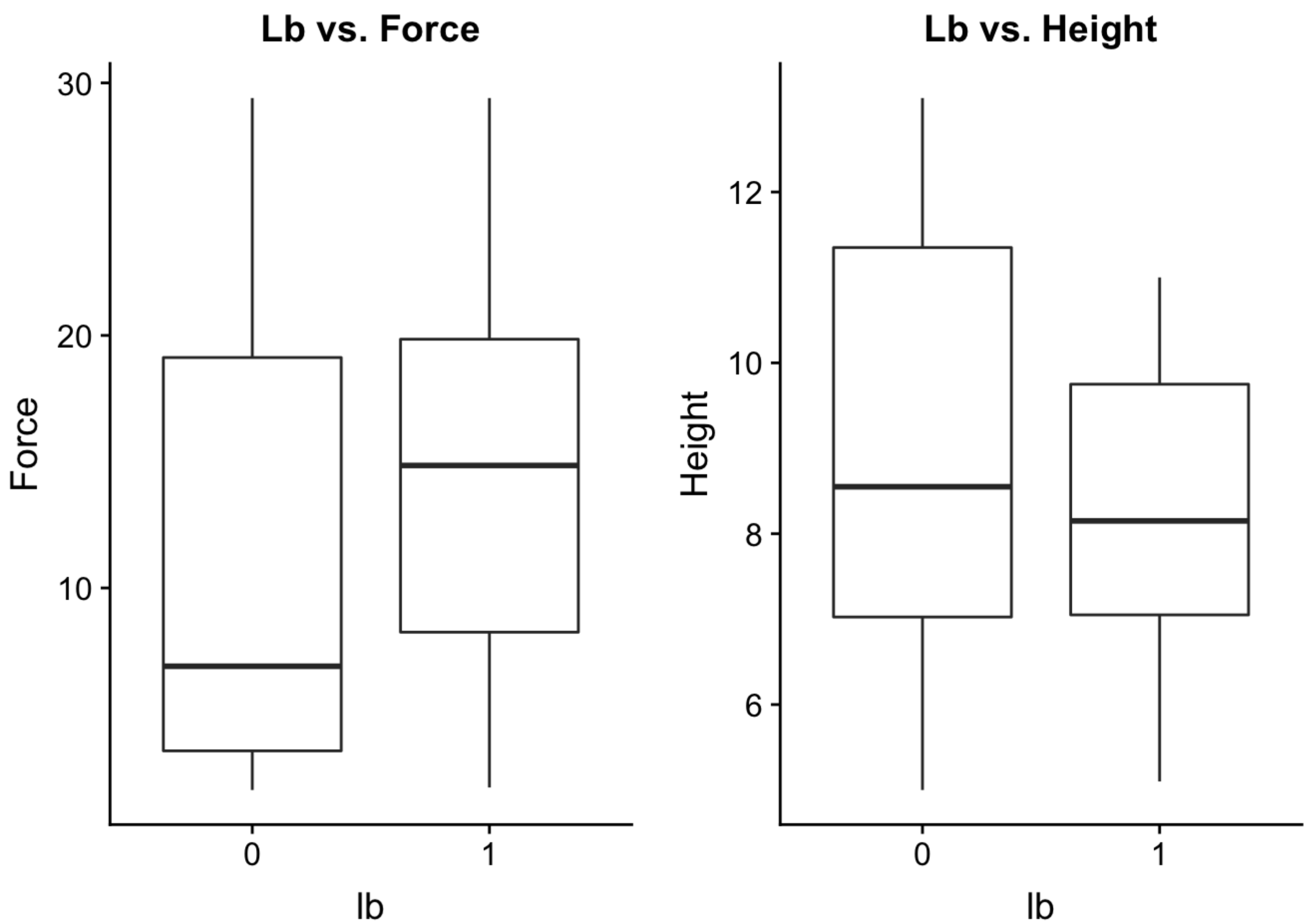
2. Briefly explain how you would conduct exploratory data analysis.

Below are some plots that can be used to conduct exploratory data analysis.

```
p1 <- ggplot(data = claws, aes(x = Height)) +
  geom_histogram() +
  labs(title = "Distribution of Height",
       x = "Height (in mm)")

p2 <- ggplot(data = claws, aes(x = Force)) +
  geom_histogram() +
  labs(title = "Distribution of Force",
       x = "Force (in Newtons)")

p3 <- ggplot(data = claws, aes(x = lb)) +
  geom_bar() +
  labs(title = "Distribution of Lb vs. Not Lb")

top_row <- plot_grid(p1, p2, ncol = 2)

plot_grid(top_row, p3, ncol = 1)
```



```
p1 <- ggplot(data = claws, aes(x = lb, y = Force)) +
  geom_boxplot() +
  labs(title = "Lb vs. Force")

p2 <- ggplot(data = claws, aes(x = lb, y = Height)) +
  geom_boxplot() +
  labs(title = "Lb vs. Height")

plot_grid(p1, p2, ncol = 2)
```



We will use the mean-centered variables for `Force` and `Height`. The model output is below.

term	estimate	std.error	statistic	p.value
(Intercept)	-1.130	0.463	-2.443	0.015
forceCent	0.211	0.092	2.279	0.023
heightCent	-0.895	0.398	-2.249	0.025

3. Write the equation for the odds of a crab being from the Lb species.

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -1.130 + 0.211 \times \text{ForceCent} - 0.895 \times \text{heightCent}$$

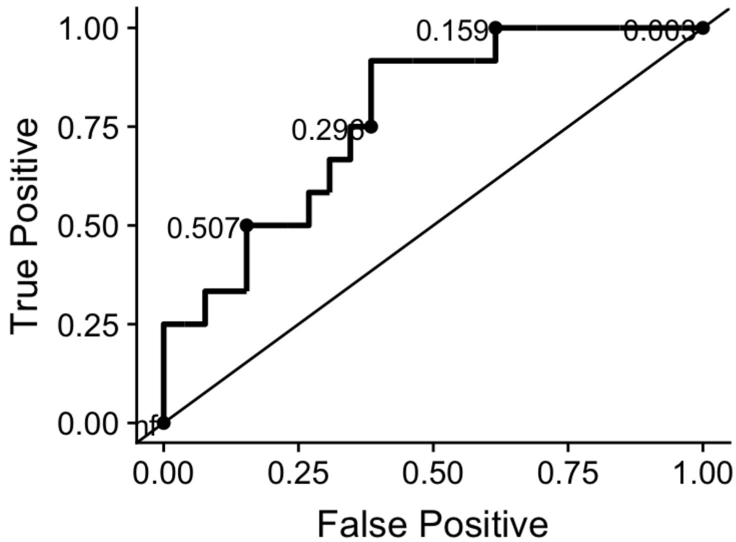
4. Interpret the intercept in the context of the problem.

The odds that a crab with average claw force (12.134) and average claw height (8.813) is the Lb species is exp(-1.130) = 0.323.

5. Interpret `forceCent` in the context of the problem.

For each additional Newton in force, we expect the odds that a crab is the Lb species to multiply by exp(0.211) = 1.235 holding height constant.

The ROC curve is below.




```
## [1] 0.775641
```

6. What does *sensitivity* mean in the context of this data? What does *specificity* mean?

Sensitivity: Correctly predicting a crab is the Lb species. Specificity: Correctly predicting a crab is not in the Lb species.

7. Suppose we use a threshold of 0.507. What is the *sensitivity* at this threshold? What is the specificity?

Sensitivity is about 0.5, and specificity is about 0.85.

Part 2: Which species?

Suppose we want to use force and height to determine a crab's species. The model output is below:

y.level	term	estimate	std.error	statistic	p.value
Hn	(Intercept)	-1.193	1.106	-1.079	0.281
Hn	forceCent	-0.494	0.196	-2.514	0.012
Hn	heightCent	0.179	0.474	0.378	0.705
Lb	(Intercept)	0.021	0.602	0.034	0.973
Lb	forceCent	0.095	0.101	0.941	0.347
Lb	heightCent	-0.902	0.429	-2.103	0.035

1. Write the equation of the model.

$$\log\left(\frac{\hat{p}_{Hn}}{\hat{p}_{Cp}}\right) = -1.193 - 0.494 \times \text{ForceCent} + 0.179 \times \text{heightCent}$$

$$\log\left(\frac{\hat{p}_{Lb}}{\hat{p}_{Cp}}\right) = 0.021 + 0.095 \times \text{ForceCent} - 0.902 \times \text{heightCent}$$

2. Interpret the intercept for the odds a crab is Hn vs. Cp species.

The odds that a crab with average claw force (12.134) and average claw height (8.813) is the Hn species vs. Cp species is exp(-1.193) = 0.303.

3. Interpret the coefficient of `forceCent` for the odds a crab is Lb vs. Cp species.

For each additional Newton in force, we expect the odds a crab is the Lb species vs. Cp species to multiply by exp(0.095) = 1.1 holding height constant.