AE 12: Exam 3 Review

Add your name here
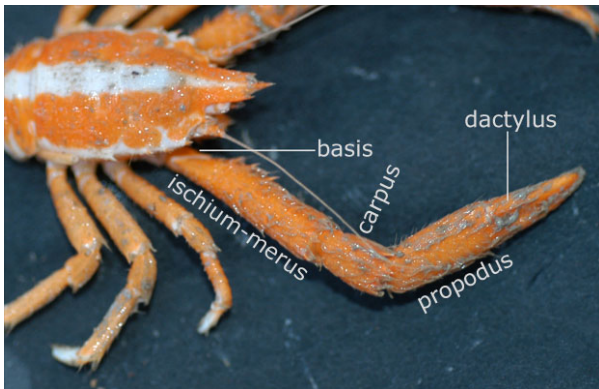
Introduction

Packages

```r
library(tidyverse)
library(tidymodels)
library(knitr)
library(Stat2Data)
library(rms)
library(nnet)
```

Data

As part of a study of the effects of predatory intertidal crab species on snail populations, researchers measured the mean closing forces and the propodus heights of the claws on several crabs of three species.

```
claws <- read_csv("data/claws.csv")
```

We will use the following variables:

- `force`: Closing force of claw (newtons)
- `height`: Propodus height (mm)
- `species`: Crab species - Cp(Cancer productus), Hn (Hemigrapsus nudus), Lb(Lophopanopeus bellus)
- `lb`: 1 if Lophopanopeus bellus species, 0 otherwise
- `hn`: 1 if Hemigrapsus nudus species, 0 otherwise
- `cp`: 1 if Cancer productus species, 0 otherwise
- `force_cent`: mean centered force
- `height_cent`: mean centered height

Before we get started, let's make the categorical and indicator variables factors.

```
claws <- claws %>%
  mutate(
    species = as_factor(species),
    lb = as_factor(lb),
    hn = as_factor(hn),
    cp = as_factor(cp)
  )
```

Part 1

Probabilities vs. odds vs. log-odds

Why we use log-odds as response variable: https://sta210-s22.github.io/website/slides/lec-18.html#/do-teenagers-get-7-hours-of-sleep

Exercise 1

Fill in the blanks:

- Use log-odds to …
- Use odds to …
- Use probabilities to …

Exercise 2

Suppose we want to use force to determine whether or not a crab is from the Lophopanopeus bellus (Lb) species. Why should we use a logistic regression model for this analysis?

Exercise 3

We will use the mean-centered variables for force in the model. The model output is below. Write the equation of the model produced by R. Don't forget to fill in the blanks for ….

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-0.798	0.358	-2.233	0.026	-1.542	-0.123
force_cent	0.043	0.039	1.090	0.276	-0.034	0.123

Let π be…

$$\log\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) =$$

Exercise 4

Interpret the intercept in the context of the data.

Exercise 5

Interpret the effect of force in the context of the data.

Exercise 6

Now let's consider adding `height_cent` to the model. Fit the model that includes `height_cent`. Then use AIC to choose the model that best fits the data.

```
lb_fit_2 <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(lb ~ force_cent + height_cent, data = claws)

tidy(lb_fit_2, conf.int = TRUE)
```

```
# A tibble: 3 x 7
  term         estimate std.error statistic p.value conf.low conf.high
  <chr>           <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
1 (Intercept)    -1.13     0.463     -2.44  0.0146   -2.17    -0.306
2 force_cent      0.211    0.0925     2.28  0.0227    0.0563    0.424
3 height_cent    -0.895    0.398     -2.25  0.0245   -1.82     -0.234
```

Exercise 7

What do the following mean in the context of this data. Explain and calculate them.

- Sensitivity: …
- Specificity: …
- Negative predictive power: …

Part 2

Let's extend the model to use force and height to predict the species (Hn, Cp, and Lb). The model output is below:

```
species_fit <- multinom_reg() %>%
  set_engine("nnet") %>%
  fit(species ~ force_cent + height_cent, data = claws)

species_fit <- repair_call(species_fit, data = claws)

tidy(species_fit, conf.int = TRUE)
```

```
# A tibble: 6 x 8
  y.level term       estimate std.error statistic p.value conf.low conf.high
  <chr>   <chr>         <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
1 Lb      (Intercept)    1.21     1.11      1.10   0.273   -0.956     3.38
2 Lb      force_cent    0.589    0.204      2.88   0.00393   0.189    0.989
3 Lb      height_cent  -1.08     0.514     -2.11   0.0352   -2.09    -0.0753
4 Cp      (Intercept)    1.19     1.10      1.08   0.281   -0.974     3.36
5 Cp      force_cent    0.494    0.197      2.51   0.0120    0.109    0.879
6 Cp      height_cent  -0.180    0.475     -0.380  0.704   -1.11     0.750
```

Exercise 8

Write the equation of the model.

$$\log\left(\frac{\hat{\pi}_{Hn}}{\hat{\pi}_{Cp}}\right) =$$

$$\log\left(\frac{\hat{\pi}_{Lb}}{\hat{\pi}_{Cp}}\right) =$$

Exercise 9

- Interpret the intercept for the odds a crab is Hn vs. Cp species.
- Interpret the effect of force on the odds a crab is Lb vs. Cp species.

Exercise 10

Interpret the effect of force on the odds a crab is in the Hn vs. Lb species.

CAUTION: We can write an interpretation based on the estimated coefficients; however, we can't make any inferential conclusions for this question based on the current model. We would need to refit the model with Lb as the baseline category to do so.
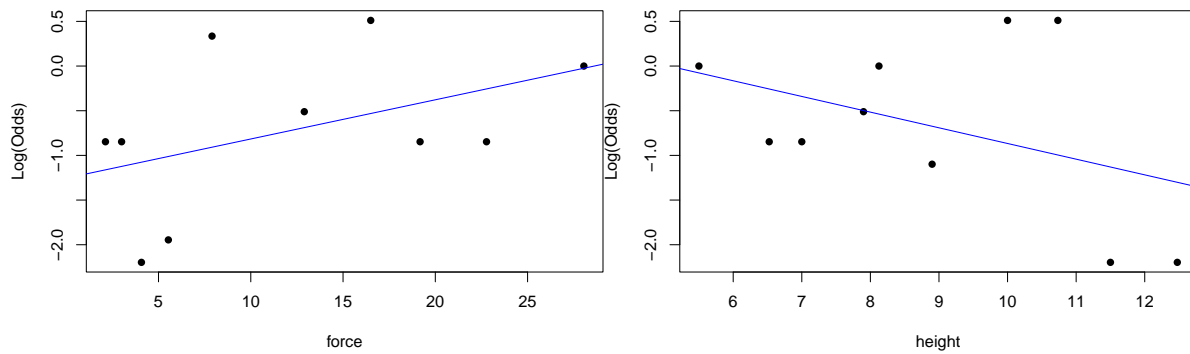
Exercise 11

Conditions for multinomial logistic (and logistic models as well):

- Independence:
- Randomness:
- Linearity:

```
emplogitplot1(lb ~ force, data = claws, ngroups = 10)
```

```
emplogitplot1(lb ~ height, data = claws, ngroups = 10)
```



```
# add code here for other species here
```

```
# add code here for other species here
```

Part 3

Checking for multicollinearity in logistic and multinomial logistic

Similar to multiple linear regression, we can also check for multicollinearity in logistic and multinomial logistic models.

- Use the `vif` function to check for multicollinearity in logistic regression.

- The `vif` function doesn't work for the multinomial logistic regression models, so we can look at a correlation matrix of the predictors as a way to assess if the predictors are highly correlated: