ST344: Report 3

u2046965

2022-11-02

Introduction

Ankle sprains are a common injury endured by individuals of all ages. The CAST (Collaborative Ankle Support Trial) compared 4 mechanical supports for acute, severe ankle sprain. This report explores a subset of the data consisting of information for 565 randomly selected patients.

This report aims to extract insight for 3 questions:

1) How does BMI vary with age and sex?
2) How does pain at baseline vary with age, BMI and sex?
3) What is an appropriate model for whether adl scores are missing at 9 months?

Data

The dataset used in this report is a subset of the data collected in the CAST (Collaborative Ankle Support Trial) study. The original data was collected from a randomised controlled trial from 8 emergency departments across the UK in order to compare four mechanical supports for acute, severe ankle sprain. Patients were sent questionnaires at months 1, 3 and 9 after their injuries.

The dataset used in this report contained the following variables:

- Weight (kg)
- Height (cm)
- Patient Age
- Patient BMI
- Baseline Pain Score (0 = the worst pain, 100 = no pain at all)
- Gender
- m9adl (whether a function in daily living response was missing or not at 9 months)

Methods

In order to answer questions 1 and 2, I created new features in the dataset titled "age_groups", "bmi_groups" and "bpain_groups" to covert the numerical variables to become categorical in order to make analysis easier. The BMI groups were made under NHS definitions. I then utilised data visualisation and summary tables to explore and answer the questions listed in the introduction.

In this report I also utilized GLMs, specifically Logistic Regression to model whether or not a patient reports function in daily living response at 9 months (m9ald). My motivation for selecting Logistic Regression is due to the binary nature of the m9adl variable and thus belonging to the binomial family.

To determine which features/predictors to include in the model, I performed both-direction stepwise regression as well as simply fitting a Logistic Regression model using all features and looking at the p-values for each feature to determine significance.

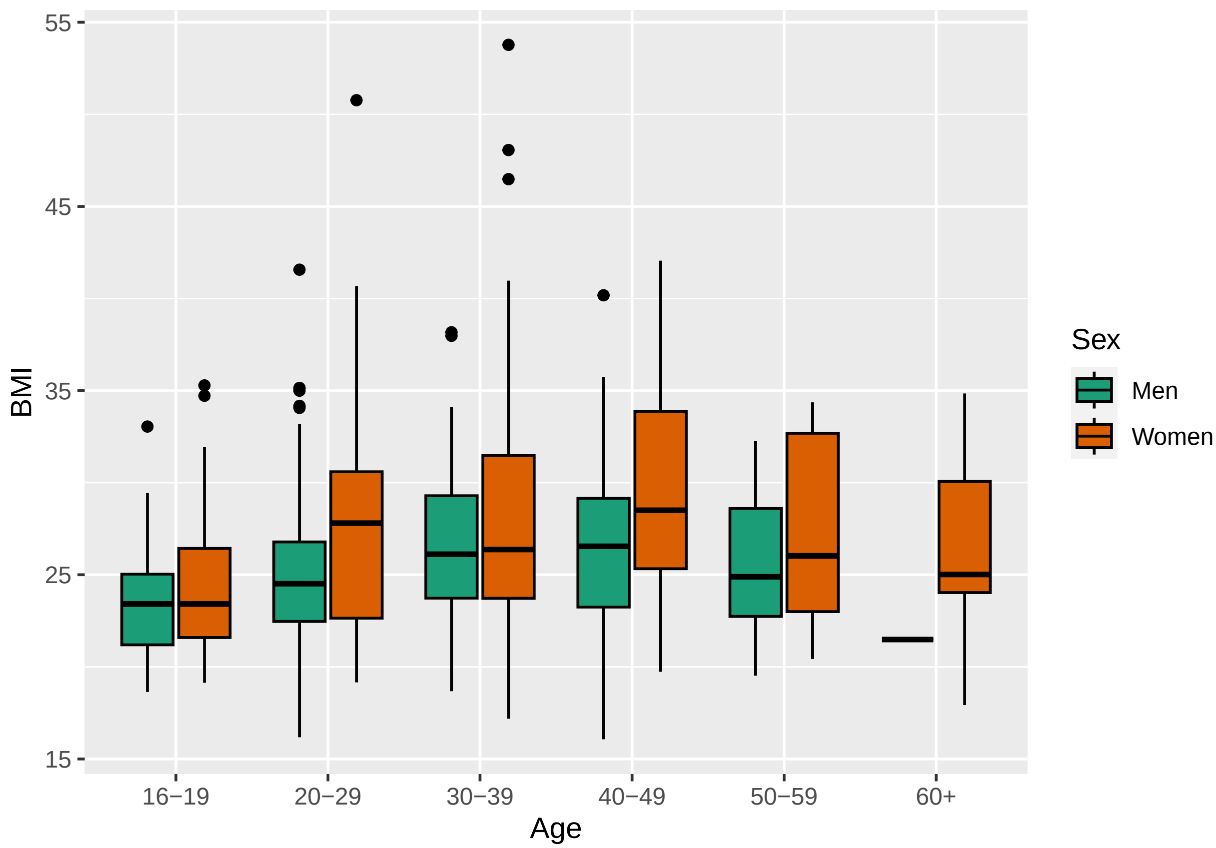
Results

Distribution of BMI by Age and Sex

The distribution of BMI by age and gender shown in figure 1 unveils key features to note. Firstly, we can see that for all age ranges, females had median BMI scores greater than or equal to the median scores for males. This is shown by the solid black line within the orange (female) boxplots being higher or level with those in the green (male) boxplots for all ages.

We can also note that, relative to their female counterparts, males had a smaller interquartile range indicating that males had less variation in BMI scores compared to females. Moreover, across all age ranges, the maximum and upper quartile BMI for a male was always less than the maximum and upper quartile BMI for a female.

Figure 1. Distribution of BMI by Age and Sex



Distribution of Baseline Pain with Age, BMI and Sex

Table 1 summaries the distribution of baseline pain against age, BMI and sex. The table is read column-by-column.

The table uncovers a few important features which should be noted. Firstly, the number of individuals with a healthy BMI reporting baseline pain values increases the higher up the score groups went before falling in the (75-100] group. Moreover, we also see that the healthy BMI individuals make up most of these injuries as, for bpain group, they make up the large majority. Obese individuals come 3rd, possibly because of an idle lifestyle compared to the more active lifestyle of their healthy peers.

Table 1: Summary of the distribution of Baseline Pain against Age, BMI and Sex

Baseline Pain	**[0-25**], N = 139	**(25-50**], N = 309	**(50-75**], N = 113	**(75-100**], N = 4
Age				
16-19	19 (14%)	57 (18%)	24 (21%)	0 (0%)
20-29	51 (37%)	118 (38%)	42 (37%)	0 (0%)
30-39	38 (27%)	72 (23%)	23 (20%)	2 (50%)
40-49	20 (14%)	51 (17%)	19 (17%)	1 (25%)
50-59	8 (5.8%)	9 (2.9%)	4 (3.5%)	1 (25%)
60+	3 (2.2%)	2 (0.6%)	1 (0.9%)	0 (0%)
BMI				
Underweight	2 (1.4%)	4 (1.3%)	0 (0%)	0 (0%)
Healthy	59 (43%)	145 (48%)	63 (57%)	1 (33%)
Overweight	45 (33%)	90 (30%)	34 (31%)	1 (33%)
Obese	32 (23%)	60 (20%)	14 (13%)	1 (33%)
Unknown	1	10	2	1
Sex				
Men	70 (50%)	186 (60%)	71 (63%)	2 (50%)
Women	69 (50%)	123 (40%)	42 (37%)	2 (50%)

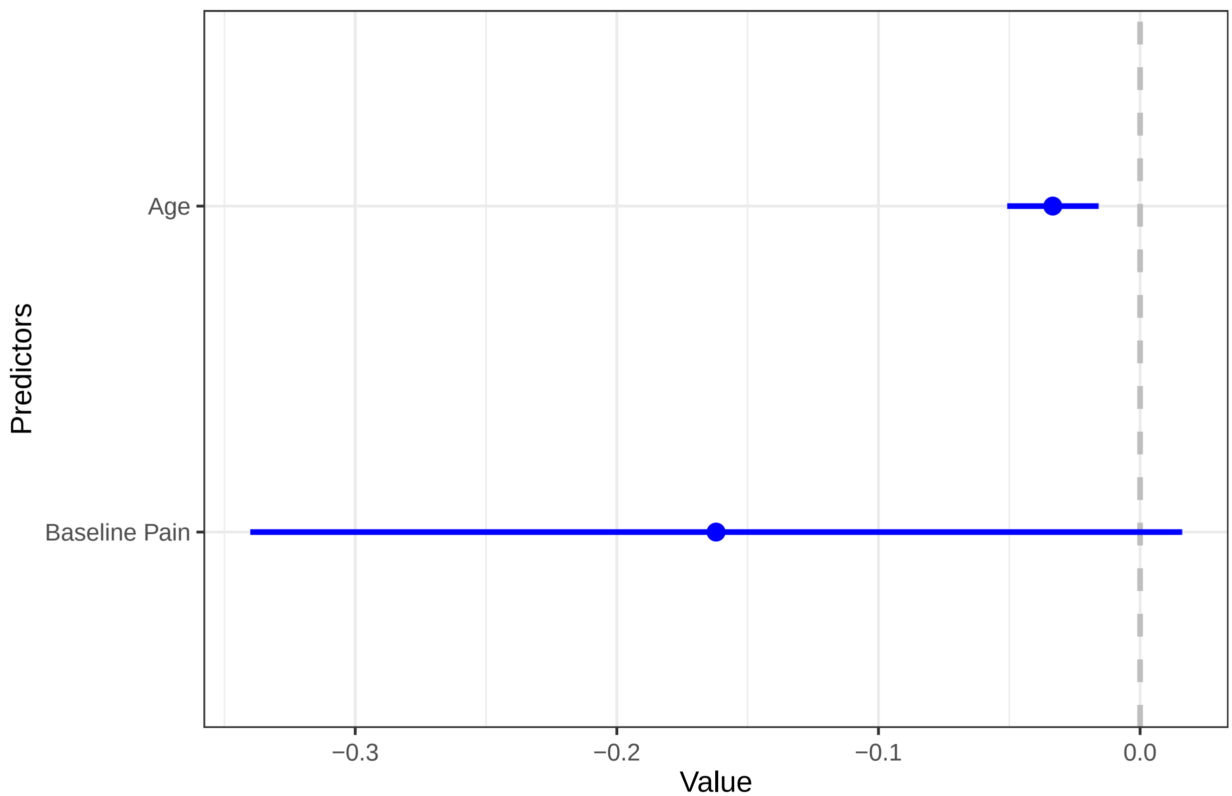
Modelling Function in Daily Living Response at 9 months

As discussed, I utilized Logistic Regression to model m9adl due to the binary nature of the variable. After performing predictor selection utilizing table 2 below, we can see that the variable age is statistically significant as it is less than 0.05 level of significance. However, after performing stepwise regression, both age and baseline pain reduced AIC.

```
##                  Estimate  Std. Error    z value     Pr(>|z|)
## (Intercept)  5.470978348 8.231932889  0.6646043 0.5063035908
## age         -0.033567919 0.009200841 -3.6483535 0.0002639263
## height      -0.025763393 0.048215491 -0.5343385 0.5931073626
## weight       0.008702512 0.051265973  0.1697522 0.8652050179
## bpain       -0.010276690 0.005483534 -1.8740998 0.0609166762
## bmi         -0.023717590 0.146961381 -0.1613865 0.8717889601
## SexWomen    -0.463997103 0.251923086 -1.8418205 0.0655014108
```

Therefore, the final model is summarised in the Figure 2 below, which plots the values of the coefficients of age and bpain and the 95% confidence interval surrounding them.

Figure 2. Logistic Regression Coefficients



Conclusions

After selecting predictors and fitting a model to the data, I was able to determine the coefficients of age and baseline pain. The value of odds ratio for bpain (standardised) can be interpreted as the odds of a response increasing by a factor of 0.85 for a bpain above 38 regardless of age compared those with a baseline score below 38.

We can also conclude females had higher BMI scores across all age groups than males, but more males had sprained their ankles than females indicating that BMI may not necessarily impact probability of ankle sprain.