LabReport3

2140365

2023-11-06

Introduction

This report contains the analysis for FAOS FOOT & ANKLE SURVEY. The Survey is used for Therapists to keep track patients foot/ankle situation and activity performance, this can help them provide a good treatment.

Data

The data, collected via questionnaires from patients with ankle injuries at 1, 3, and 9 months, encompasses 6 variables: age, sex, height, weight, bfaosFDL (baseline function in daily living score), and miss9sport (indicating sports performance deterioration). Surveying 560 participants, including 235 females and 325 males, ages ranged from 16 to 72 years (mean age: 29.98 years). Among males, heights ranged from 160.02 cm to 200.66 cm, and among females, from 147.32 cm to 187.96 cm. The average height was 172.89 cm. Weights varied between 39.92 kg and 133.36 kg, with an overall average of 78.64 kg. 214 people of all missed sport performance after 9 months. See Table1.

Table 1: Summary Table for FAOS FOOT ANKLE SURVEY

	Female	Male	All
Total number of patients			
NumPatients	235.00	325.00	560.00
Age			
MaxAge	72.00	60.00	72.00
MinAge	16.00	16.00	16.00
MeanAge	32.68	28.02	29.98
Height			
MaxHeight	187.96	200.66	200.66
MinHeight	147.32	160.02	147.32
MeanHeight	164.88	178.69	172.89
VariancHeight	52.03	48.78	96.52
Weight			
MaxWeight	133.36	127.01	133.36
MinWeight	39.92	50.80	39.92
MeanWeight	75.48	80.92	78.64
VarianceWeight	270.95	204.66	239.26
bfaosFDL			
MaxbfasFDL	88.24	100.00	100.00
MinbfasFDL	0.00	0.00	0.00
MeanbfasFDL	56.78	58.22	57.61
VariancebfasFDL	189.34	199.72	195.52
Num of people miss sport for 9months			
Num	81.00	133.00	214.00

Method

what method is used to get the result of analysis: glm, glm.nb The model used in analysis, model1,2,3,description of models, what it is looks like, why use this model(final model,with step function, describe DF, residual deviance, p-value)

To discover How does weight vary with age, sex and height, a generalised log-linear model is used here. can be written as

$$log(weight) = \beta_0 1 + \beta_2 sex + \beta_3 height + \beta_4 age + \epsilon$$

Data will be fitted better when logarithm is applied to the explanatory variable. Notice that, here sex is considered as factor, it will consider data with respect to different gender.

To know Does function in daily living at baseline vary with age, sex, weight or height, $cor()$ function is used to find out the correlation between bfasFDL and other 4 variables.

To get the appropriate model, $step()$ function in R is applied,which is used find out which model would be the most appropriate model and Deviance and AIC value will be given in each step; 5 models are built to get the best one:

- Model1: $miss9sport \sim weight + height + age + sex + bfaosFDL$
- Model2: $miss9sport \sim height + age + sex + bfaosFDL$
- Model3: $miss9sport \sim height + age + sex$
- Model4: $miss9sport \sim age + sex$
- Model5: $miss9sport \sim age$

Result

From methods, the result for How does weight vary with age, sex and height can be explainedas following. These coefficients for sex0 and sex1(2.504051 and 2.454716) indicate that, holding other variables constant, individuals of these two genders have a higher body weight compared to the reference gender. For instance, the average weight of sex1 will be higher than that of sex0 (when other factors remain constant), so that sex1 refers to males and sex0 refers to females.

For height, the coefficient for height is 0.010262. This means that holding other variables constant, for every one-unit increase in height, weight will increase by 1.010315kg. Therefore, individuals with taller stature tend to have a higher body weight.

Same for age. The coefficient for age is 0.003208. This signifies that holding other variables constant, for every one-unit increase in age, the weight will increase by 1.003213kg. This implies a slight upward trend in weight with increasing age.

From function of $cor()$, the result shows that there is only slightly correlation between bfaosFDL and age, height and weight, -0.086,0.031 and -0.048 respectively.

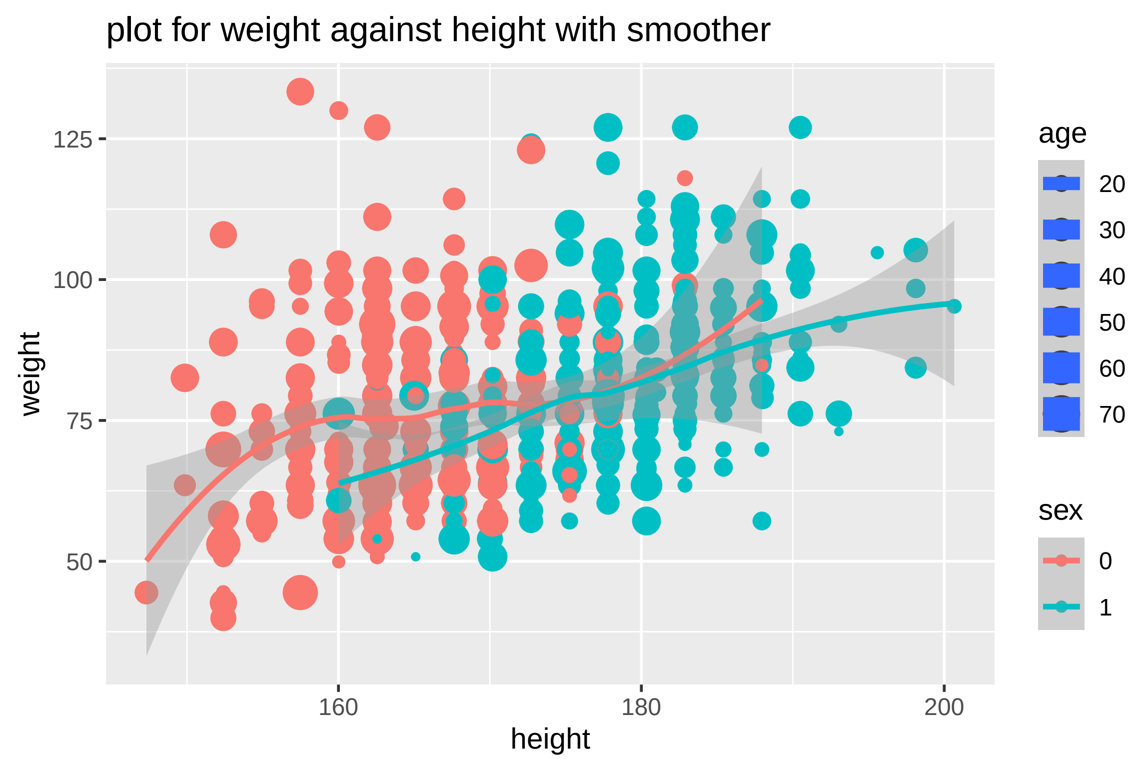
The most appropriate model is $miss9sport \sim age$, this is got from the $step()$ function from start from Model1. The formal form is,

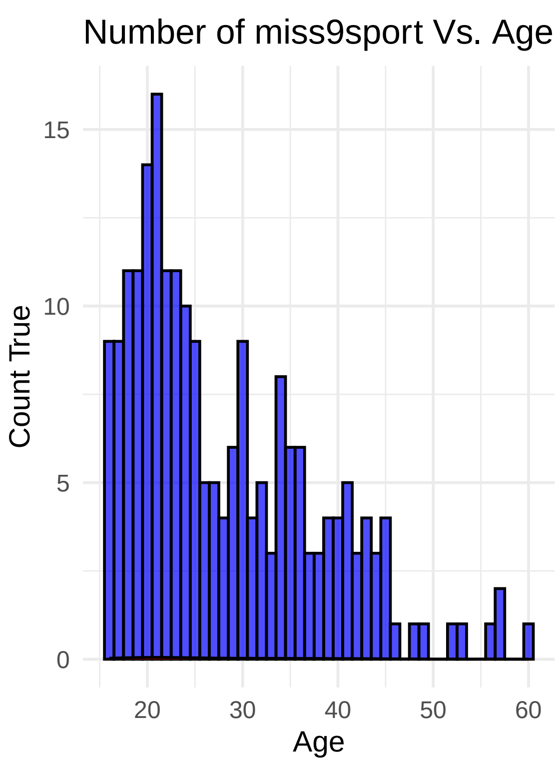
$$miss9sport = \beta_0 age + \epsilon$$

$\beta_0$here is -0.01742, can be explained as: with 1 year older, there is 1.7% to miss the sport performance. This result can also get from figure2, number of people miss9sport and number of people not miss9sport against Age.

The summary table for 5 models shows the reason of why model5 is the most appropriate model. "Age" is always have p-value < 0.15, so it is contained in model. Then model4 have intercept have p-value < 0.15, but p-value for sex is too big. Same reason for other models, large p-value is always contained in model which is not aprropriate. Therefore $model \sim age$ is the most appropriate model.

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```


plot for weight against height with smoother
weight
125
100
75
50
160
180
200
height
age
20
30
40
50
60
70
sex
0
1


Number of miss9sport Vs. Age
Count True
15
10
5
0
20
30
40
50
60
Age

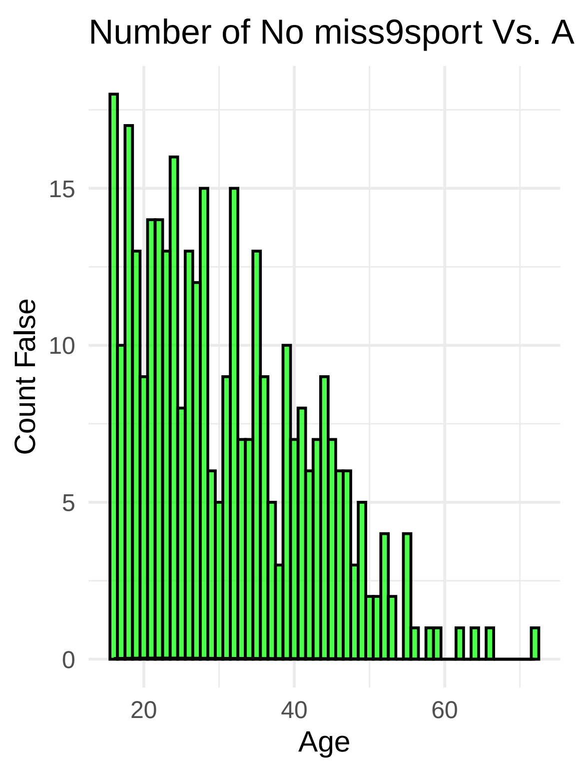

Number of No miss9sport Vs. A
Count False
15
10
5
0
20
40
60
Age

Table 2: Summary Table for 5 models

	Residual_DF	Residual_deviance	AIC_value	p_value
Model1				
intercept1	554	403.01	843.03	0.4297
weight1	554	403.01	843.03	0.9087
height1	554	403.01	843.03	0.3199
sex1	554	403.01	843.03	0.2191
age1	554	403.01	843.03	0.0132
bfaosFDL1	554	403.01	843.03	0.5287
Model2				
intercept2	555	403.02	841.04	0.4344
height2	555	403.02	841.04	0.3059
sex2	555	403.02	841.04	0.2208
age2	555	403.02	841.04	0.0119
bfaosFDL2	555	403.02	841.04	0.5256
Model3				
intercept3	556	404.07	839.44	0.3250
sex3	556	404.07	839.44	0.4680
age3	556	404.07	839.44	0.0170
bfaosFDL3	556	404.07	839.44	0.5460
Model4				
intercept4	557	404.43	838.45	0.0236
sex4	557	404.43	838.45	0.4754
age4	557	404.43	838.45	0.0236
Model5				
intercept5	558	404.95	836.96	0.0260
age5	558	404.95	836.96	0.0114

Conclusions

From data provided, there is relationship between all variables. - weight gets larger when either age or height increase, and indeed, weight for male is larger than weight for female - There is correlation between bfaosFDL and age,height and weight, negative correlation with both age and height, positive correlation with weight. However, the correlation between these variable is very small. - The most appropriate model for whether sports scores are missing at nine months is $miss9sport \sim age$, this result got from $step()$ function, which get p-value, residuals and AIC value step by step to help identify the most appropriate model.

Limitations

For finding the correlation, there might be other methods to find the correlation instead use $cor()$.Also, I did not fully present the correlation between sex and bfaosFDL.