

Group 5 Analysis

1 Introduction

2 Exploratory Analysis

We begin with creating a new binary response variable based on the BMIgroup variable for obesity classification, which is divided into “Obesity” and “Not Obesity”. Then, we draw a barplot Figure 1 to show how the prevalence of obesity in Scotland changes over years. Here we can see that both obesity and non-obesity barely changed between 2013 and 2016. The proportion of obesity was the highest in 2015 and the lowest in 2014, but the difference was not significant, with only a 0.4% fluctuation.

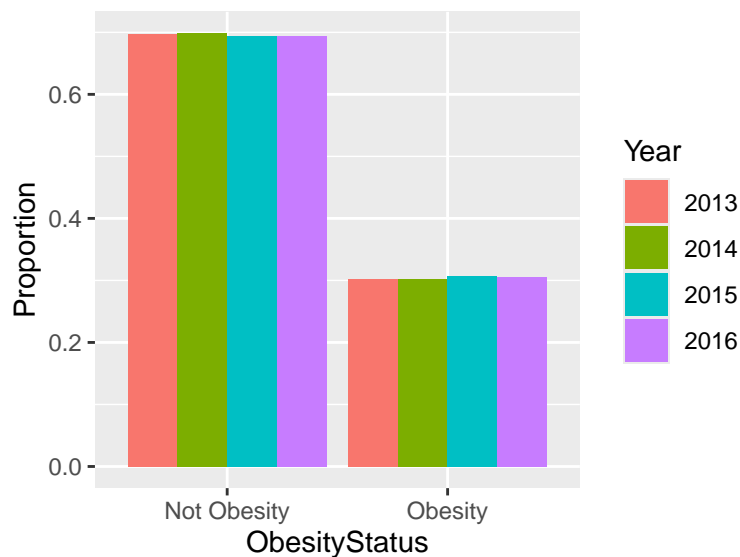


Figure 1: Barplot of Obesity by Year

To further explore whether the prevalence of obesity differs by age, gender, and employment factors, we draw different barplots on obesity by age group, sex, and employment, respectively.

Figure 2 displays the proportion of obesity in different age groups. The 16-24 age group has the lowest obesity proportion, while the 65-74 age group has the highest proportion. We can see that the proportion of obesity gradually increases from the 16-24 age group to the 65-74 age group, but the obesity prevalence among the elderly over 75 years old group is lower than that in the previous age group.

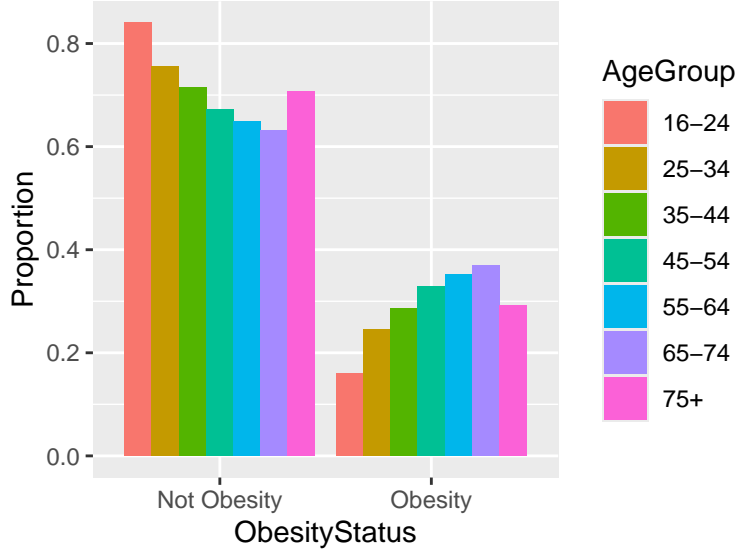


Figure 2: Barplot of Obesity by AgeGroup

Figure 3 illustrates the obesity proportion in females and males, respectively. There does not appear to be a significant difference in obesity rates between females and males, although the prevalence among females is slightly higher than that among males.

Figure 4 displays that people with full-time education have the lowest prevalence of obesity, while people who are permanently unable to work have the highest proportion. It seems likely that employment increases the prevalence of obesity, but a person is more likely to be obese if they are permanently unable to find a job.

3 Formal Analysis

We fit the logistic regression model for obesity prevalence from 2013 to 2016,

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_{\text{Year}} \cdot \mathbb{I}_{\text{Year}}(x)$$

where

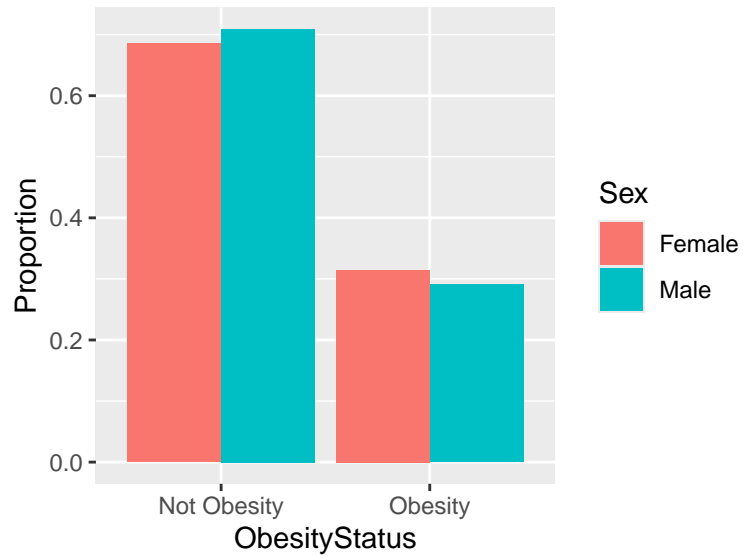


Figure 3: Barplot of Obesity by Sex

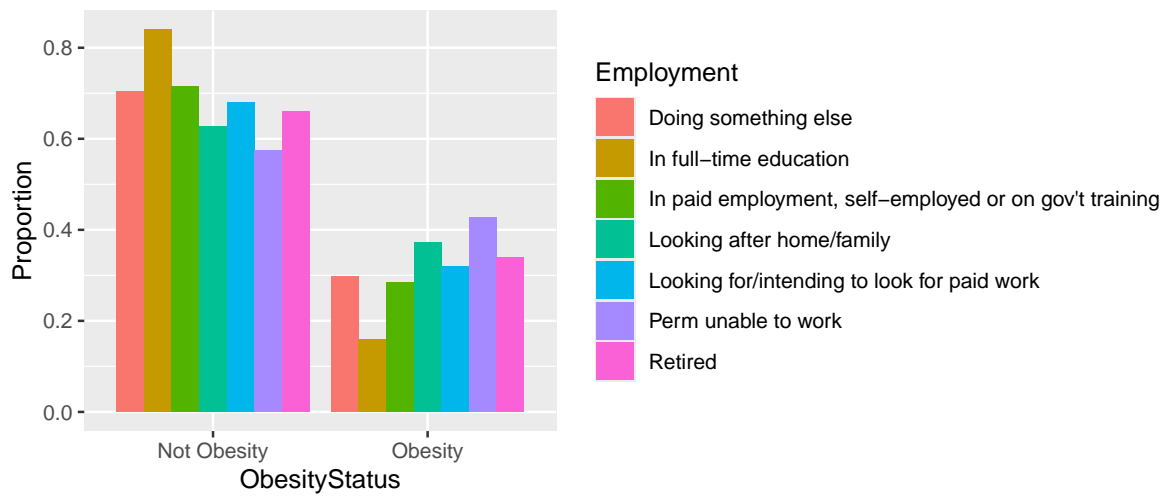


Figure 4: Barplot of Obesity by Employment

- p is the probability of obesity
- $1 - p$ is the probability of not obesity
- α is the intercept of the regression line for the baseline Year (2013)
- β_{Year} the additional term added to α to get the intercept of the regression line for the specified Year
- $\mathbb{I}_{\text{Year}}(x)$ is an indicator function indicating the chosen Year.

Table 1: Estimates of the intercept and slope from the fitted logistic regression model

term	estimate	std.error	p.value
(Intercept)	−0.8330	0.0358	0.0000
Year2014	−0.0056	0.0513	0.9125
Year2015	0.0163	0.0509	0.7493
Year2016	0.0111	0.0525	0.8318

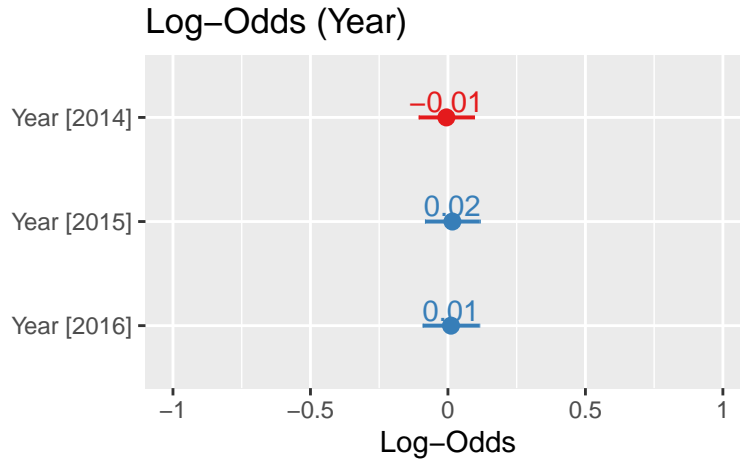


Figure 5: The log-odds for Obesity by Year

Then we fit the logistic regression model for the explanatory variables, age group, sex, and employment,

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta_{\text{AgeGroup}} \cdot \mathbb{I}_{\text{AgeGroup}}(x) + \beta_{\text{Sex}} \cdot \mathbb{I}_{\text{Sex}}(x) + \beta_{\text{Employment}} \cdot \mathbb{I}_{\text{Employment}}(x)$$

where

- p is the probability of obesity
- $1 - p$ is the probability of not obesity
- α is the intercept of the regression line for the baseline AgeGroup (16-24), Sex (Female), Employment (Doing something else)
- β_{AgeGroup} the additional term added to α to get the intercept of the regression line for the specified AgeGroup
- β_{Sex} the additional term added to α to get the intercept of the regression line for the specified Sex
- $\beta_{\text{Employment}}$ the additional term added to α to get the intercept of the regression line for the specified Employment
- $\mathbb{I}_{\text{AgeGroup}}(x)$ is an indicator function indicating the chosen AgeGroup.
- $\mathbb{I}_{\text{Sex}}(x)$ is an indicator function indicating the chosen Sex.
- $\mathbb{I}_{\text{Employment}}(x)$ is an indicator function indicating the chosen Employment.

Table 2: Coefficients from the fitted logistic regression model with all explanatory variables

term	estimate	std.error	p.value
(Intercept)	-1.5249	0.1590	0.0000
AgeGroup25-34	0.4216	0.1049	0.0001
AgeGroup35-44	0.6303	0.1029	0.0000
AgeGroup45-54	0.8157	0.1012	0.0000
AgeGroup55-64	0.9087	0.1034	0.0000
AgeGroup65-74	1.0140	0.1212	0.0000
AgeGroup75+	0.6638	0.1324	0.0000
SexMale	-0.1001	0.0379	0.0083
EmploymentIn full-time education	-0.3010	0.1741	0.0840
EmploymentIn paid employment, self-employed or on gov't training	-0.0330	0.1339	0.8054
EmploymentLooking after home/family	0.4104	0.1650	0.0129
EmploymentLooking for/intending to look for paid work	0.2626	0.1658	0.1132
EmploymentPerm unable to work	0.4734	0.1547	0.0022
EmploymentRetired	0.0198	0.1470	0.8929

4 Conclusion

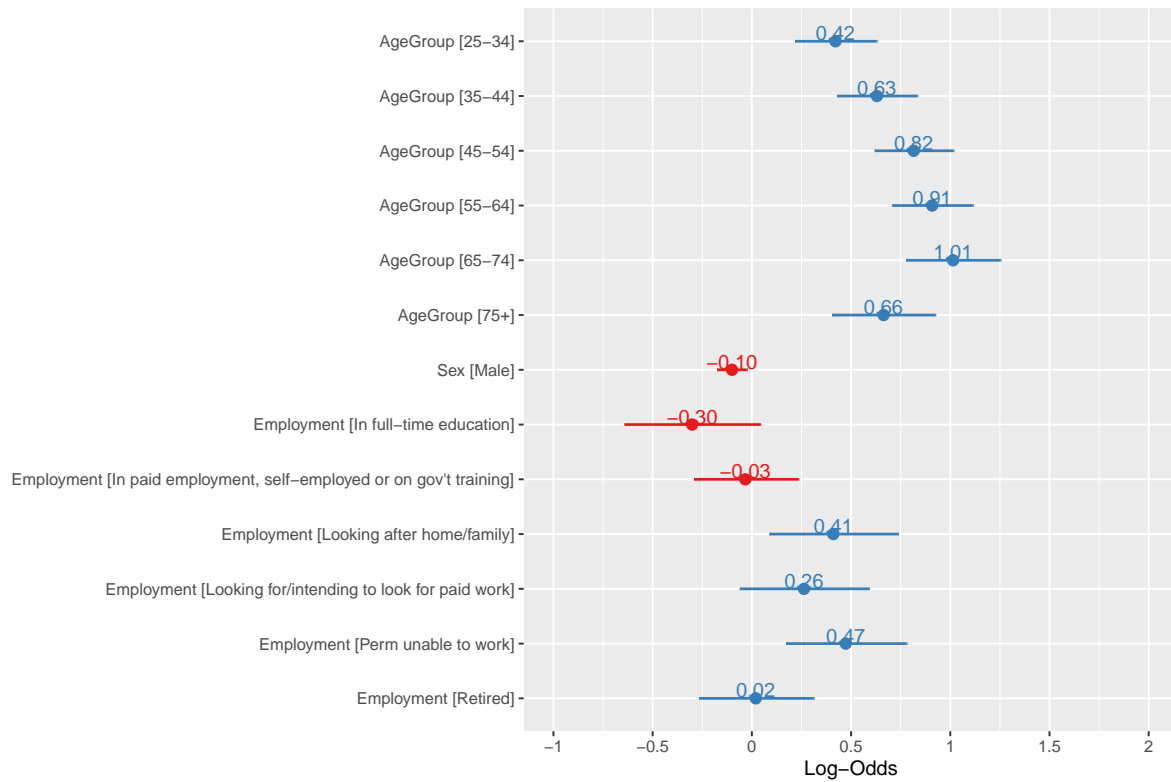


Figure 6: Log-Odds for age group, sex, employment