Has Scottish Obesity Prevalence Changed over the Years and do Different Factors Effect this

Jo Scott-Stephen, Eleanor Mills and Sophie Brown

1 Introduction

Obesity rates have been observed to be increasing in many developed countries. However it has been linked to numerous health issues, specifically cardiovascular issues, diabetes and cancer. With data supplied by the Scottish Health Survey, we will be examining the trends in obesity of the Scottish population in the years 2013-2016, to see if Scotland is following that trend.

Furthermore, we will investigate whether any specific socio-economic characteristics and lifestyle factors indicate a higher probability of obesity, to see, more generally, if there is truly a correlation between the way we live and our health.

First we will undergo exploratory data analysis in Section 2 before conducting our formal analysis in Section 3 and eventually reaching a conclusion and a discussion in Section 4.

2 Exploratory Analysis

We begin by conducting some exploratory data analysis to investigate whether obesity rates have changed over the years.

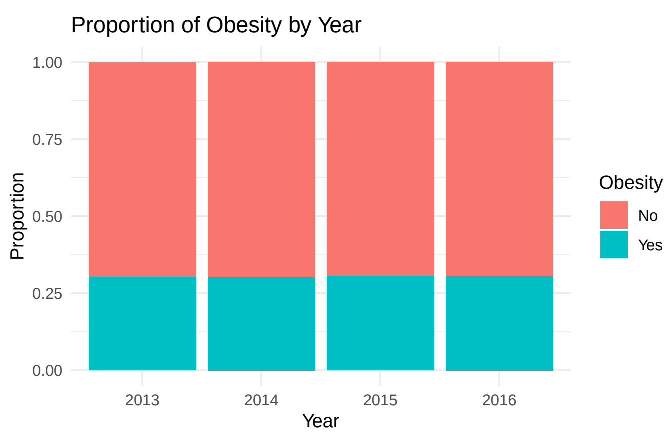

Figure 1: Proportion of Obesity by Year

Figure 1 displays the proportion of Obesity by the each year the questionnaire has been taken. There appears to be very little if any difference in the change of obesity levels over the years.

We then investigate the effect of certain lifestyle factors and socioeconomic status on obesity rates

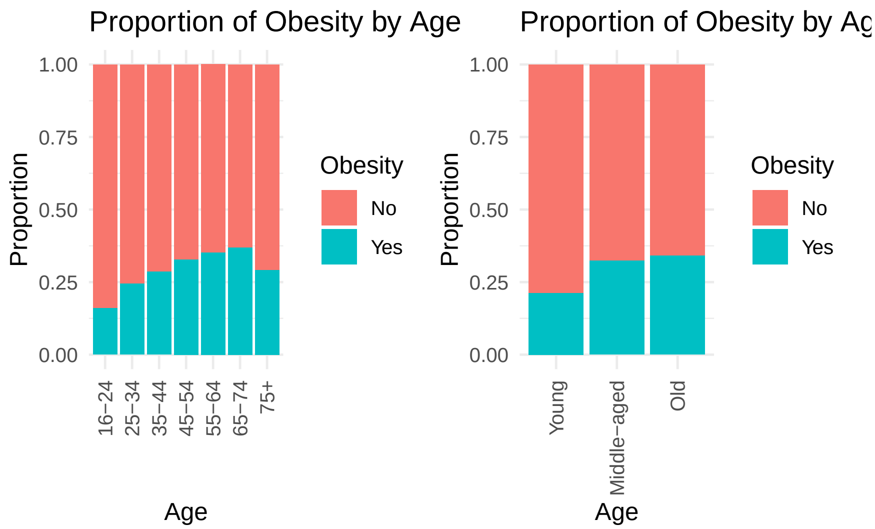

Figure 2: Obesity Proportions against Age (left) and Age Categories (right)

We first look at Figure 2 to see if there is a significant relationship between obesity and age, i.e. can the age of a person change their likelihood of being obese. The plot does indicate an increase in obesity rate as age increases however it then the rate drops in the last category (75+). The difference in ages is made clearer in the right-hand plot where the age catgegories have been refined, and we see that young people are least likely to be obese. Middle aged are much more likely to be obese but the elderly are the most likely to be obese.

We then see if there is a difference in obesity dependent on the gender of a person with Figure 3 . The plot shows there is a larger proportion of women who are obese, suggesting that if someone's sex is female they may be more likely to be obese than men.

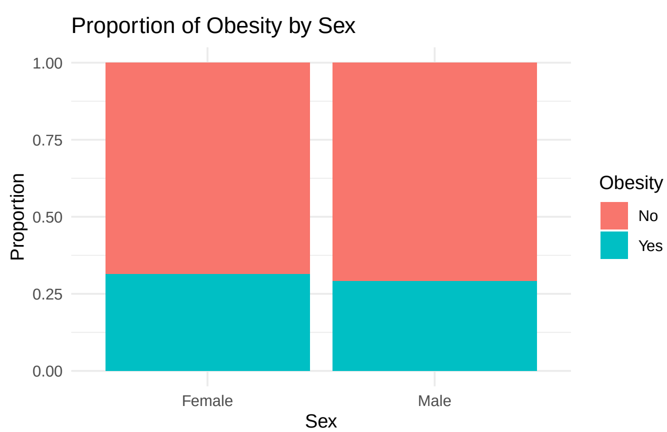

Figure 3: Proportion of Obesity by Gender

Then, using Figure 4 we invesitugae the impact of employment status on obesity rates. We have refined employment status into three categories, 'employed/full time education', 'retired', 'unemployed/other'. We can see that the least amount of obesity occurs in the employed/full time education category, while the category with highest obesity rates is the unemployed/other category which is interesting as from the previous graphs looking at the impact of age, we would have expected that those who are retired to have the highest obesity rates, however perhaps the slight drop-off that we observed for those over 75 has effected this.

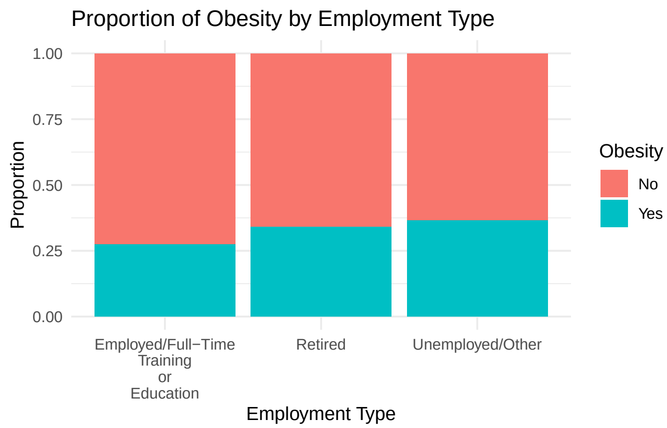

Figure 4: Proportion of Obesity by Employment Type

We then investigate the effect of people's dietary choices. Figure 5 displays a bar plot of Obesity rates vs if surveyed people ate the correct amount of fruit and vegetables in a day (lifestyle factors). We might have expected this to show there is a significant impact on obesity rates if people ate their '5 a day' however there seems to be no difference in obesity depending on a persons fruit and veg intake.

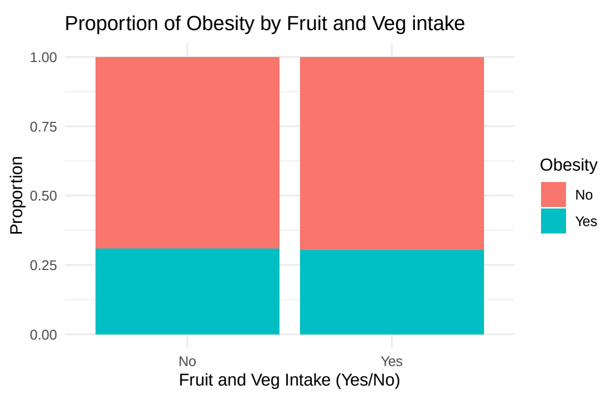

Figure 5: The Proportion of Obesity Against Consumption of the Required Amount of Fruit or Vegetables

3 Formal Analysis

We first fit the full regression model containing all explanatory variables to investigate whether obesity rate has chaanged over the years, before refining our model with modelling selection criteria based on the AIC. Since our response variable is binary, we fit a logistic regression model without interaction as we want to focus onthe individual effects of each of our covariates. The model can be written as:

$$Obese \sim Bernoulli(p_i)$$

$$logit(p_i) = \alpha + \beta_{\text{sex}} \cdot \mathbb{1}_{\text{sex}} + \beta_{\text{year}} \cdot \text{Year} + \beta_{\text{fv}} \cdot \mathbb{1}_{\text{FV}} + \beta_{\text{middle-aged}} \cdot \mathbb{1}_{\text{middle-aged}} + \beta_{\text{old}} \cdot \mathbb{1}_{\text{old}} + \beta_{retired} \cdot \mathbb{1}_{\text{retired}} + \beta_{unemployed} \cdot \mathbb{1}_{\text{unemployed}}$$

where p_i is the probability of being obese and

- α is the intercept (the baseline log-odds of obesity when all predictors are at their reference category),
- β_{sex} is the coefficient for sex,
- $\mathbb{1}_{\text{sex}}(x)$ is an indicator function such that

$$\mathbb{1}_{\text{sex}}(x) = \begin{cases} 1 & \text{if gender of } x \text{ is male,} \\ 0 & \text{if gender of } x \text{ is female.} \end{cases}$$

- β_{year} is the coefficient for survey year,
- β_{fv} is the coefficient for whether an individual consumes the daily intake of either fruit, vegetables or both,
- $\mathbb{1}_{\text{FV}}(x)$ is an indicator function such that

$$\mathbb{1}_{\text{FV}}(x) = \begin{cases} 1 & \text{if } x \text{ consumes the recommended daily intake of fruit and/or vegetables,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\beta_{\text{middle-aged}}$ is the coefficient for being middle-aged (35-64 years old),
- $\mathbb{1}_{\text{middle-aged}}(x)$ is an indicator function such that

$$\mathbb{1}_{\text{middle-aged}}(x) = \begin{cases} 1 & \text{if } x \text{ is middle-aged (35-64 years old),} \\ 0 & \text{otherwise.} \end{cases}$$

- β_{old} is the coefficient for being old (65+ years),
- $\mathbb{1}_{\text{old}}(x)$ is an indicator function such that

$$\mathbb{1}_{\text{old}}(x) = \begin{cases} 1 & \text{if } x \text{ is old (65+ years old),} \\ 0 & \text{otherwise.} \end{cases}$$

- β_{retired} is the coefficient for being retired,
- $\mathbb{1}_{\text{retired}}(x)$ is an indicator function such that

$$\mathbb{1}_{\text{retired}}(x) = \begin{cases} 1 & \text{if } x \text{ is retired,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\beta_{\text{unemployed}}$ is the coefficient for being unemployed or in another non-working status,
- $\mathbb{1}_{\text{unemployed}}(x)$ is an indicator function such that

$$\mathbb{1}_{\text{unemployed}}(x) = \begin{cases} 1 & \text{if } x \text{ is unemployed or in another non-working status,} \\ 0 & \text{otherwise.} \end{cases}$$

As we have many covariates and a rather large model we will check the significance of our covariates in Table 1 to see if we have room to refine it.

Table 1: Table looking at significance of full model covariates

Term	p-value
Intercept	0.9592
Sex: Male	0.0056
Year	0.9289
Eating Recommended Fruit and Veg: Yes	0.2416
Age Category: Middle-aged	0.0000
Age Category: Old	0.0000
Employment Status: Retired	0.1528
Employment Status: Unemployed/Other	0.0000

From Table 1 we see that many of our covariates do not have significant p-values, including the variable Year. This means that there is not enough evidence to indicate that obesity rates have changed over the years.

The AIC for this model is 17011.99, As this value is rather high and we have many insignificant covariates, we employ stepwise regression with backward selection to see if we can achieve a better fit, based on the lowest AIC. The new model we fit is defined as follows:

$$logit(p_i) = \alpha + \beta_{\text{sex}} \cdot \mathbb{1}_{\text{sex}} + \beta_{\text{middle-aged}} \cdot \mathbb{1}_{\text{middle-aged}} + \beta_{\text{old}} \cdot \mathbb{1}_{\text{old}} + \beta_{retired} \cdot \mathbb{1}_{\text{retired}} + \beta_{unemployed} \cdot \mathbb{1}_{\text{unemployed}}$$

The model has coefficients and p-values seen in the following table (Table 2)

Table 2: Estimates of the fitted model's coefficients to 4 d.p.

Term	Estimate	p-value
Intercept	-1.3402	0.0000
Male	-0.1014	0.0069
Middle-aged	0.5658	0.0000
Old	0.6249	0.0000
Retired	0.1036	0.1551
Unemployed / Other	0.4006	0.0000

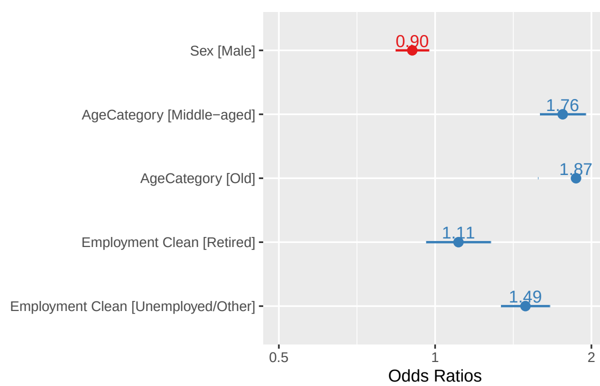

Figure 6: Odds (Obesity)

From Table 2 we see that all the covariates apart from being retired are significant, meaning sex, age and being unemployed (not retired) all effect obesity rates. In particular age is seen to be extremely significant. Being classified as middle-aged causes an increased log-odds of 0.5658 while being classified as old causes an increased log-odds of 0.631 of being obese compared to young people. There is also a significant increase in the log-odds of obesity of 0.4006 associated with being unemployed. Furthermore, we have a significant decrease in log-odds of being obese of 0.1014 associated with being male rather than female. All of which agrees with our previous plots.

These results can be visualised in Figure 6 , which visualises our odds ratios. We note that the figure agrees with our previous conclusion that being retired does not significantly impact the likelihood that someone is obese.

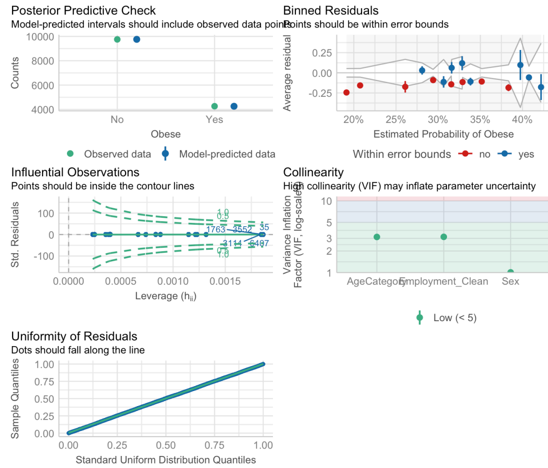

Figure 7: Model Diagnostic Plots

We then check the model fit with Figure 7 and also see how accurately it predicts obesity with Figure 8.

Figure 7 allows us to assess if our model assumptions are met. From the plots it seems that our model predicted intervals include the observed data points and none of these points are deemed influential. Furthermore, our residuals seem fairly uniform. However, over half our binned residuals lie outwith the error bounds, indicating we have an ill-fitted model.

This is further confirmed as when we check the AIC of our reduced model it decreases by a negligible amount, from 17011 to 17009. Therefore, our previous conclusions may not be appropriate.

We still go on to check how well the model predicts obesity with Figure 8

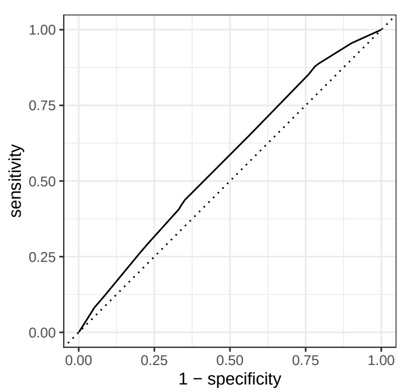

Figure 8: ROC curve for Stepwise Regression Model

The ROC curve seen in Figure 8 is fairly close to the diagonal line then this indicates that our model is only slightly better than random guessing, again indicating a poor model is fitted.

4 Conclusions

In conclusion, it seems that overall obesity rates have not changed over the years 2013-16. However, this only looks at the population as a whole, and perhaps certain subgroups of the population dependent on age or socioeconomic status have experienced some sort of change in levels of obesity over these 3 years, especially as certain characteristics may impact the likelihood that someone is obese. There seems to be some suggestion that being unemployed along with getting older as well as being female increase someone's likelihood of being obese. Interestingly, it does not seem that eating recommended amounts of fruit and vegetables significantly effect obesity rates. This could be as we looked at the effects of eating recommended amounts either fruit or vegetables, and perhaps these variables should be considered individually, especially as fruit generally has a higher sugar content. Furthermore, being retired does not have a significant effect either, which is odd as one would think that being unemployed and retired would yield a similar lifestyle.

However, as our model assumptions have been violated these conclusions may not be appropriate and should be investigated more thoroughly, especially as our ROC curve performed poorly, indicating our model predicts obesity not much better than random chance. Moreover, further research should be conducted into how age effects obesity rates with age being considered as a discrete variable instead of a categorical one especially because the left plot seen in Figure 2 seems to indicate there exists a slight decrease in obesity rates for those over 75. This could be due to increased hospital visits or perhaps just generally poorer health. This also somewhat agrees with the finding that retirement does not significantly impact obesity levels. However, as our findings are incredibly limited due to our model assumptions being violated, another model should be fitted, perhaps with implementing some sort of transformation or perhaps with a different model altogether.