Case Study 3 - Group 1

Annika Janson h11829506

Jan Beck h11814291

10.5.2021

# 1. 1

### 1.1.1

### 1.1.2

### 1.1.3

### 1.1.4

# 1.2

### 1.2.1

Linear Probability Model

## [1] FALSE

## [1] TRUE

##   
## Call:  
## lm(formula = low ~ smoke + race + age + lwt + ptl + ht + ui +   
## ftv, data = babies)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.8398 -0.3154 -0.1413 0.4040 0.9299   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.510256 0.209531 2.435 0.01586 \*   
## smoke 0.159886 0.071293 2.243 0.02614 \*   
## race2 0.221511 0.100423 2.206 0.02867 \*   
## race3 0.144410 0.076831 1.880 0.06179 .   
## age -0.003705 0.006441 -0.575 0.56590   
## lwt -0.002545 0.001162 -2.190 0.02982 \*   
## ptl 0.115828 0.068276 1.696 0.09154 .   
## ht 0.366239 0.135466 2.704 0.00752 \*\*  
## ui 0.156567 0.092992 1.684 0.09399 .   
## ftv 0.006343 0.031115 0.204 0.83868   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4354 on 179 degrees of freedom  
## Multiple R-squared: 0.1637, Adjusted R-squared: 0.1217   
## F-statistic: 3.894 on 9 and 179 DF, p-value: 0.0001557

Linear Probit Model

##   
## Call:  
## glm(formula = low ~ smoke + race + age + lwt + ptl + ht + ui +   
## ftv, family = binomial(link = "probit"), data = babies)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8845 -0.8269 -0.5217 0.9904 2.2444   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.270956 0.700911 0.387 0.69907   
## smoke 0.569099 0.234676 2.425 0.01531 \*   
## race2 0.749313 0.314295 2.384 0.01712 \*   
## race3 0.522137 0.255541 2.043 0.04103 \*   
## age -0.018469 0.021669 -0.852 0.39405   
## lwt -0.008906 0.003995 -2.229 0.02579 \*   
## ptl 0.319749 0.208341 1.535 0.12485   
## ht 1.111117 0.416607 2.667 0.00765 \*\*  
## ui 0.465138 0.279296 1.665 0.09583 .   
## ftv 0.028439 0.101609 0.280 0.77956   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 234.67 on 188 degrees of freedom  
## Residual deviance: 201.04 on 179 degrees of freedom  
## AIC: 221.04  
##   
## Number of Fisher Scoring iterations: 5

Linear Logit Model

##   
## Call:  
## glm(formula = low ~ smoke + race + age + lwt + ptl + ht + ui +   
## ftv, family = binomial(link = "logit"), data = babies)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8943 -0.8209 -0.5319 0.9820 2.2124   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.477788 1.196800 0.399 0.68973   
## smoke 0.938848 0.402105 2.335 0.01955 \*   
## race2 1.271730 0.527306 2.412 0.01588 \*   
## race3 0.880960 0.440722 1.999 0.04562 \*   
## age -0.029585 0.037029 -0.799 0.42429   
## lwt -0.015395 0.006918 -2.225 0.02605 \*   
## ptl 0.543415 0.345384 1.573 0.11563   
## ht 1.862276 0.697428 2.670 0.00758 \*\*  
## ui 0.767576 0.459301 1.671 0.09469 .   
## ftv 0.065463 0.172384 0.380 0.70413   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 234.67 on 188 degrees of freedom  
## Residual deviance: 201.31 on 179 degrees of freedom  
## AIC: 221.31  
##   
## Number of Fisher Scoring iterations: 4

### 1.2.2

AIC und BIC

## [1] 233.8057 221.0447 221.3051

## [1] 269.4649 253.4621 253.7226

Looking at AIC the Linear Logit Model is the best. This is supported by the results of the BIC as in both cases the Linear Logit Model has the smallest numbers. The numbers of AIC and BIC of the Probit Model only slightly higher to the numbers of the Linear Logit Model tho, so it seems they are nearly similar good.

### 1.2.3

The results of 1.2.1 show that in the Linear Probabiity Model, the variables smoke, race2(=black), lwt, ht. In the Linear Probit Model the same values are significant, but race 3 (= others) is joining. The Linear Logit Models is similar to the Probit Model, which is inline with the results of BIP and AIC, that they are similar in explaining the response variable **low** and that they are explaining it better than the linear Probability Model. It seems that smoking has an important effect, as it is present in all three models. It makes sense that the weight of a mother influences the weight of a baby.

### 1.2.4

Predictions:

## 1   
## 0.05547055

## 1   
## 0.07036926

## 1   
## 0.08060108

The highest probability predicted for that a baby has a birth weight under 2500kg with given variables is predicted by the Linear Logit Model. With a probability of **0.08060108**, so around **8.06%** it is likely that a child is born with under 2500kg with these circumstances. As we observed before the Linear Probit Model fits a little bit better, so we want take a look at the predicted probabilty in this model. It is **0.07036926**, so around **7.03%** and around 1% lower than the probability of the Linear Logit Model.

#2 Theory