Lab 8

Grace Randall

2024-03-19

Question 1

```
fit_1 <- glm(disease ~ 1, family="binomial",data=disease_df)
summary(fit_1)
```

```
##
## Call:
## glm(formula = disease ~ 1, family = "binomial", data = disease_df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.3888     0.1289   3.017  0.00256 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 337.3  on 249  degrees of freedom
## Residual deviance: 337.3  on 249  degrees of freedom
## AIC: 339.3
##
## Number of Fisher Scoring iterations: 4
```

a) the value of beta 0 is 0.388. This is the log odds of having disease.
b) the predicted probability of finding the disease is 0.59.

Question 2

```
fit_2 <- glm(disease ~ old, family="binomial",data=disease_df)
summary(fit_2)
```

```
##
## Call:
## glm(formula = disease ~ old, family = "binomial", data = disease_df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.7053     0.1909   3.695  0.00022 ***
## old          -0.6100     0.2613  -2.335  0.01956 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 337.30  on 249  degrees of freedom
## Residual deviance: 331.78  on 248  degrees of freedom
## AIC: 335.78
##
## Number of Fisher Scoring iterations: 4
```

a) the estimate for beta 0 is 0.705. this is very different from what it was in the previous model. This is the log odds of having disease in a new growth forest.
b) The probability of the disease in old growth forests is predicted to be 0.534.
c) The probability of the disease in new growth forests is predicted to be 0.669.
d) the odds ratio for old growth forest compared to new growth forests is 0.543. this means that the odds of finding disease in an old growth forest are 0.543 times the odds of finding disease in new growth forests.
e) We can reject the hypothesis of no difference between the old and new growth forests with a p-value of 0.019.

Question 3

```
mean_ba=mean(disease_df$basal_area)
disease_df <- mutate(disease_df, basal_area_centered=basal_area-mean_ba)

fit_3 <- glm(disease ~ old + basal_area_centered, family="binomial",data=disease_df)
summary(fit_3)
```

```
##
## Call:
## glm(formula = disease ~ old + basal_area_centered, family = "binomial",
##     data = disease_df)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          7.053e-01  1.916e-01   3.682 0.000232 ***
## old                 -6.101e-01  2.632e-01  -2.318 0.020467 *
## basal_area_centered  7.968e-05  2.660e-02   0.003 0.997610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 337.30  on 249  degrees of freedom
## Residual deviance: 331.78  on 247  degrees of freedom
## AIC: 337.78
##
## Number of Fisher Scoring iterations: 4
```

a) the odds ratio for basal area is 1.0000796. This means that an increase in basal area of 1m2 will result in the probability of disease being multiplied by 1.0000796.
b) We fail to reject the null hypothesis of no effect of basal area with a p-value of 0.997.

Question 4

```
fit_4 <- glm(disease ~ old + basal_area + temp_anomaly, family="binomial",data=disease_df)
summary(fit_4)
```

```
##
## Call:
## glm(formula = disease ~ old + basal_area + temp_anomaly, family = "binomial",
##     data = disease_df)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.133693   0.783146   0.171  0.86445
## old          -0.575481   0.268812  -2.141  0.03229 *
## basal_area   -0.005912   0.027329  -0.216  0.82873
## temp_anomaly  0.481414   0.154128   3.123  0.00179 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 337.30  on 249  degrees of freedom
## Residual deviance: 321.66  on 246  degrees of freedom
## AIC: 329.66
##
## Number of Fisher Scoring iterations: 4
```

a) the odds ratio for temperature anomalies is 1.618. An increase in the temperature anomaly of 1 results in the odds of disease being multiplied by 1.618.

b) If I were to be creating a model myself I would absolutely keep in the old vs new growth aspect and the temperature anomoly aspect because both of these are both statistically significant and have a major effect on the odds. I would likely also keep in the basal area because while it does not cause major changes over 1 m2 of change, basal area might have a greater effect over the full range of possible values. I may consider removing it though because it sounds a lot more difficult to measuere so a model may be more applicable if it can predict a little less of the variation but could be applied more due to lower requirements for data availability.