final question 3

2024-04-24

Question 3, part (a)

```
data <- read.csv("/Users/sophiebuer/Downloads/ulcer.csv")

# Convert factors to factor type
data$case_control <- factor(data$case_control, levels = c("control", "case"))
data$ulcer <- factor(data$ulcer, levels = c("duodenal", "gastric"))
data$aspirin <- factor(data$aspirin, levels = c("non-user", "user"))

# Model 1: case_control and ulcer are jointly independent of aspirin
model1 <- glm(case_control ~ ulcer * aspirin, data = data, family = binomial)

# Model 2: case_control and aspirin are conditionally independent given ulcer type
model2 <- glm(case_control ~ ulcer + aspirin, data = data, family = binomial)

# Model 3: No three-factor interaction model
model3 <- glm(case_control ~ ulcer + aspirin + ulcer:aspirin, data = data, family = binomial)

summary(model1)
```

```
##
## Call:
## glm(formula = case_control ~ ulcer * aspirin, family = binomial,
##     data = data)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8
## -1.177  -1.177   1.177   1.177  -1.177  -1.177   1.177   1.177
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                7.850e-16  1.414e+00       0        1
## ulcergastric              -1.570e-15  2.000e+00       0        1
## aspirinuser               -1.256e-15  2.000e+00       0        1
## ulcergastric:aspirinuser   2.512e-15  2.828e+00       0        1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11.09  on 7  degrees of freedom
## Residual deviance: 11.09  on 4  degrees of freedom
## AIC: 19.09
##
## Number of Fisher Scoring iterations: 2
```

```
summary(model2)
```

```
##
## Call:
## glm(formula = case_control ~ ulcer + aspirin, family = binomial,
##     data = data)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8
## -1.177  -1.177   1.177   1.177  -1.177  -1.177   1.177   1.177
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.355e-16  1.225e+00       0        1
## ulcergastric  1.570e-16  1.414e+00       0        1
## aspirinuser   3.140e-16  1.414e+00       0        1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11.09  on 7  degrees of freedom
## Residual deviance: 11.09  on 5  degrees of freedom
## AIC: 17.09
##
## Number of Fisher Scoring iterations: 2
```

```
summary(model3)
```

```
##
## Call:
## glm(formula = case_control ~ ulcer + aspirin + ulcer:aspirin,
##     family = binomial, data = data)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8
## -1.177  -1.177   1.177   1.177  -1.177  -1.177   1.177   1.177
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                7.850e-16  1.414e+00       0        1
## ulcergastric              -1.570e-15  2.000e+00       0        1
## aspirinuser               -1.256e-15  2.000e+00       0        1
## ulcergastric:aspirinuser  2.512e-15  2.828e+00       0        1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11.09  on 7  degrees of freedom
## Residual deviance: 11.09  on 4  degrees of freedom
## AIC: 19.09
##
## Number of Fisher Scoring iterations: 2
```

Question 3, part (b)

```r
odds_ratios <- exp(coef(model1))
odds_ratios
```

```
##              (Intercept)             ulcergastric              aspirinuser
##                        1                        1                        1
## ulcergastric:aspirinuser
##                        1
```

Looking at the odds ratios from the joint independence model, we can see that coeficients for factors 'ulcergastric' and 'aspirinuser' and their interaction 'ulcergastric:aspirinuser' are all exactly 1, which means that there's no association between these factors and the case-control status when taking into account the interaction between the two factors. I guess that means that ulcer type doesn't affect relationship between aspririn use and being case or control and vice versa.

Question 3, part (c)

```r
anova(model1, model3, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: case_control ~ ulcer * aspirin
## Model 2: case_control ~ ulcer + aspirin + ulcer:aspirin
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         4      11.09
## 2         4      11.09  0        0
```

The joint independence model (model1) is adequate and there is no evidence to suggest that the more complex model with interaction term (model3) is a better fit than model1.

Question 3, part (d) Model1 is the best model out of the three models in (a) and is as good as the saturated model (model3). Since model3 and model1 are just as good as each other and model1 is simpler, model1 is best.

Question 3, part (e)

```r
marginal_probability <- mean(data$case_control == "case" & data$ulcer == "gastric" & data$aspirin == "u
marginal_probability
```

```
## [1] 0.125
```

The marginal probability that someone is a case, has a gastric ulcer, and is an aspirin user is 12.5%.