# Appendix

## 1. Preparation

library(here)

## here() starts at /Users/amalinaismail/Desktop/DrPH/ SEM 2/ADVANCED CATEGORICAL/ASSIGNMENT

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(haven)  
library(gtsummary)  
library(VGAM)

## Loading required package: stats4  
## Loading required package: splines

library(nnet)  
library(broom)  
library(knitr)  
library(kableExtra)

##   
## Attaching package: 'kableExtra'  
##   
## The following object is masked from 'package:dplyr':  
##   
## group\_rows

library(tibble)  
library(purrr)  
library(gt)  
library(ggplot2)  
library(ggeffects)  
library(reshape2)

##   
## Attaching package: 'reshape2'  
##   
## The following object is masked from 'package:tidyr':  
##   
## smiths

library(data.table)

##   
## Attaching package: 'data.table'  
##   
## The following objects are masked from 'package:reshape2':  
##   
## dcast, melt  
##   
## The following objects are masked from 'package:lubridate':  
##   
## hour, isoweek, mday, minute, month, quarter, second, wday, week,  
## yday, year  
##   
## The following objects are masked from 'package:dplyr':  
##   
## between, first, last  
##   
## The following object is masked from 'package:purrr':  
##   
## transpose

library(ordinal)

##   
## Attaching package: 'ordinal'  
##   
## The following objects are masked from 'package:VGAM':  
##   
## dgumbel, dlgamma, pgumbel, plgamma, qgumbel, rgumbel, wine  
##   
## The following object is masked from 'package:dplyr':  
##   
## slice

library(foreign)  
library(gridExtra)

##   
## Attaching package: 'gridExtra'  
##   
## The following object is masked from 'package:dplyr':  
##   
## combine

library(grid)  
library(viridis)

## Loading required package: viridisLite

library(ggpubr)  
library(rmarkdown)  
library(webshot)  
library(webshot2)

## Registered S3 method overwritten by 'webshot2':  
## method from   
## print.webshot webshot  
##   
## Attaching package: 'webshot2'  
##   
## The following objects are masked from 'package:webshot':  
##   
## appshot, resize, rmdshot, shrink, webshot

## 2. Read data

datafh <- read\_csv("fetal\_health.csv")

## Rows: 2126 Columns: 22  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (22): baseline value, accelerations, fetal\_movement, uterine\_contraction...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

summary(datafh)

## 3. Convert fetal health to factor

datafh$fetal\_health <- as.factor(datafh$fetal\_health)  
summary(datafh)

## 4. Data Wrangling

datafh <- datafh %>%   
 select(fetal\_health, `baseline value`, abnormal\_short\_term\_variability , uterine\_contractions, fetal\_movement)

glimpse(datafh)

### 4a. Rename Column

datafh <- datafh %>%  
 rename(  
 fetalhealth = fetal\_health,  
 baseline = `baseline value`,  
 abnvar = abnormal\_short\_term\_variability ,  
 uterinecontractions = uterine\_contractions,  
 fetalmovement = fetal\_movement  
 )

## 5. Create new categorical variable from fetal\_health

datafh %>% group\_by(fetalhealth) %>% count()

datafh <-   
 datafh %>%   
 mutate(fetalhealth2 = factor(fetalhealth,   
 labels = c("normal","suspect","pathological")))  
datafh %>%   
 count(fetalhealth2)

## 6. Exploratory data analysis

datafh %>%  
 select(-fetalhealth) %>%  
 tbl\_summary(by = fetalhealth2,  
 statistic = list(all\_continuous() ~ "{mean} ({sd})"))

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

### Plots

#### Fetal health Distribution

* # Recreate fetal\_health\_counts with the new variable names  
  fetal\_health\_counts <- datafh %>% count(fetalhealth)  
  fetal\_health\_counts$percent <- round(fetal\_health\_counts$n / sum(fetal\_health\_counts$n) \* 100, 2)  
    
  # Define the custom colors  
  custom\_colors <- c('#ffb6c1', '#ffffe0', '#ffcc99') # Light pink, light yellow, light orange  
    
  # Pie chart for fetalhealth  
  p1 <- ggplot(fetal\_health\_counts, aes(x = "", y = n, fill = factor(fetalhealth))) +  
   geom\_bar(width = 1, stat = "identity") +  
   coord\_polar("y") +  
   geom\_text(aes(label = paste0(percent, "%")), position = position\_stack(vjust = 0.5)) +  
   labs(title = "Fetal Health", fill = "Fetal Health") +  
   scale\_fill\_manual(values = custom\_colors) +  
   theme\_void()  
    
  # Bar chart for fetalhealth  
  p2 <- ggplot(fetal\_health\_counts, aes(x = factor(fetalhealth), y = n, fill = factor(fetalhealth))) +  
   geom\_bar(stat = "identity") +  
   geom\_text(aes(label = n), vjust = -0.5) +  
   labs(title = "Fetal Health", x = "Fetal Health", y = "Count") +  
   scale\_fill\_manual(values = custom\_colors) +  
   theme\_minimal()  
    
  # Arrange plots in a grid  
  grid.arrange(p1, p2, ncol = 2, top = textGrob("Fetal Health Distribution", gp = gpar(fontsize = 20, fontface = "bold")))

#### Baseline

* datafh |>  
   ggplot(aes(baseline)) +   
   geom\_histogram() +   
   facet\_grid(. ~ fetalhealth2)
* ## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

#### Abnormal short term variability

* datafh |>  
   ggplot(aes(abnvar)) +   
   geom\_histogram() +   
   facet\_grid(. ~ fetalhealth2)
* ## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

#### Uterine contractions

* datafh |>  
   ggplot(aes(uterinecontractions)) +   
   geom\_histogram() +   
   facet\_grid(. ~ fetalhealth2)
* ## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

#### Fetal movements

* datafh |>  
   ggplot(aes(fetalmovement)) +   
   geom\_histogram() +   
   facet\_grid(. ~ fetalhealth2)
* ## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Confirm the order

levels(datafh$fetalhealth2)

### Pathological as smallest order

datafh <- datafh %>%   
 mutate(fetalhealth2 = fct\_relevel(fetalhealth2,   
 c("pathological", 'suspect', 'normal')))  
levels(datafh $fetalhealth2)

# 7. Estimation

## 1.VGAM

## 7a. Single independent variable

#### 1. Baseline

log\_baseline <- vglm(fetalhealth2 ~ baseline,   
 multinomial, data = datafh)  
summary(log\_baseline)

#### 2. Abnormal short term variability

* log\_abnvar <- vglm(fetalhealth2 ~ abnvar,   
   multinomial, data = datafh)  
  summary(log\_abnvar)

#### 3. Uterine contractions

* log\_utecon <- vglm(fetalhealth2 ~ uterinecontractions,   
   multinomial, data = datafh)

#### 4. Fetal movement

* log\_fetmov <- vglm(fetalhealth2 ~ fetalmovement,   
   multinomial, data = datafh)  
  summary(log\_fetmov)

## 

## 7b. Multiple independent variable

mlog <- vglm(fetalhealth2 ~ baseline + abnvar + uterinecontractions + fetalmovement,   
 multinomial, data = datafh)  
summary(mlog)

## 7c. Model with interaction term between independent variables

mlogi <- vglm(fetalhealth2 ~ baseline + abnvar + uterinecontractions + fetalmovement + uterinecontractions\*fetalmovement,   
 multinomial, data = datafh)

summary(mlogi)

# 2. NNET Package

mlog\_nnet <- multinom(fetalhealth2 ~ baseline, data = datafh)

summary(mlog\_nnet)

datafh <- datafh %>%

mutate(fetalhealth2\_relev = relevel(fetalhealth2, ref = "normal"))

levels(datafh$fetalhealth2\_relev)

mlog\_nnet\_rel <- multinom(fetalhealth2\_relev~ baseline, data = datafh )

# 8. Inferences

## VGAM package

### Log odds

b\_mlog <- coef(mlog)  
ci\_mlog <- confint(mlog)   
b\_ci\_mlog <- data.frame(b\_mlog,ci\_mlog) %>%  
 rename("log odds" = b\_mlog, "Lower CI" = X2.5.., "Upper CI" = X97.5..)  
b\_ci\_mlog %>%   
 kbl(digits = 2, booktabs = T, caption = "Log odds from multinomial logistic regression") %>%  
 kable\_styling(position = "center")

### Relative Risk Ratio

rrr\_mlog <- exp(b\_ci\_mlog)  
tab\_mlog <- cbind(b\_ci\_mlog, rrr\_mlog)  
colnames(tab\_mlog) <- c('b', 'lower b', 'upper b',  
 'RRR', 'lower RRR', 'upper RRR')  
tab\_mlog %>%  
 kbl(digits = 2, booktabs = T, caption = "Log odds and RRR from multinomial logistic regression") %>%  
 kable\_styling(position = "center")

# NNet package

z.test <- summary(mlog\_nnet\_rel1)$coefficients/summary(mlog\_nnet\_rel1)$standard.errors

# 2-tailed

p.val <- (1 - pnorm(abs(z.test), 0, 1)) \* 4

colnames(p.val) <- c('p-val intercept', 'p-val baseline', 'p-val abnvar', 'p-val uterinecontractions', 'p-val fetalmovement')

p.val

confint(mlog\_nnet\_rel1, level=0.95)

# 9. Predictions

## Predict the Log odds

summary(log\_baseline)

Predict the first 6 observations

head(predict.vgam(log\_baseline, type = 'link'))

Verify

head(datafh)[1:3,]

## Predict the probability

head(predict.vgam(log\_baseline, type = 'response'))

## Plot the probability

# Plotting the data ---------------------------  
data\_plot <- datafh %>%  
 ggplot(aes(x = baseline, y = fetalhealth2, color = fetalhealth2))+  
 geom\_point()+  
 coord\_cartesian(xlim = c(min(datafh$baseline), max(datafh$baseline)))  
  
  
# Predicting the probabilities ---------------  
new\_data <- data.frame(baseline = seq(min(datafh$baseline), max(datafh$baseline), by = 0.1))  
prediction <- as.data.frame(predict(log\_baseline, new\_data, type = 'response'))  
new\_data <- cbind(new\_data, prediction)  
  
# Plotting the probabilities -----------------  
prob\_plot <- new\_data %>%  
 pivot\_longer(-baseline, names\_to = "fetalhealth2", values\_to = "probability") %>%  
 ggplot(aes(x = baseline, y = probability, color = fetalhealth2))+  
 geom\_line()  
  
# Merging the two plots -----------------------  
ggarrange(data\_plot, prob\_plot, nrow = 2)