# final question 4

2024-04-24

**Question 4, part (a)**

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

# proportion of dead fetuses
data$dead_proportion <- data$R / data$N

# binomial GLM with canonical logit link
model_a <- glm(dead_proportion ~ grp + hb, family = binomial, data = data, weights = data$N)

summary(model_a)
```

```
##
## Call:
## glm(formula = dead_proportion ~ grp + hb, family = binomial,
##     data = data, weights = data$N)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -4.601  -1.013  -0.117   1.469   2.755
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.02696    0.37759   5.368 7.96e-08 ***
## grpTreatment -2.65088    0.48238  -5.495 3.90e-08 ***
## hb           -0.18713    0.07428  -2.519   0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 509.43  on 57  degrees of freedom
## Residual deviance: 170.57  on 55  degrees of freedom
## AIC: 248.04
##
## Number of Fisher Scoring iterations: 5
```

The coefficients for 'grpTreamtment' and 'hb' are both statistically significant since $p<0.05$. The coefficient for 'grpTreatment' suggests that treatment is associated with a decrease by 2.65 in log odds of proportion of dead fetuses per litter. The coefficient for 'hb' suggests that a higher baseline hemoglobin levels are associated with a -0.187 decrease in log odds of fetal deaths.

**Question 4, part (b)**

```r
model_b <- glm(dead_proportion ~ grp + hb, family = binomial(link = "log"), data = data, weights = data$

summary(model_b)
```

```
##
## Call:
## glm(formula = dead_proportion ~ grp + hb, family = binomial(link = "log"),
##     data = data, weights = data$N)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.7041  -1.1466  -0.0467   1.3622   2.9886
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.09939    0.12884   0.771  0.44047
## grpTreatment -1.89885    0.28883  -6.574 4.89e-11 ***
## hb           -0.08149    0.02908  -2.802  0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 509.43  on 57  degrees of freedom
## Residual deviance: 167.70  on 55  degrees of freedom
## AIC: 245.16
##
## Number of Fisher Scoring iterations: 7
```

The coefficient for both 'grpTreatment' and 'hb' are statistically significant suggesting that the treatment and baseline hemoglobin levels affect the natural log of the proportion of dead fetuses. Being in the treatment group decreases the natural log of proportion of dead fetuses by 1.89885. Higher baseline hemoglobin levels decreases natural log of proportion of dead fetuses by 0.08149.

**Question 4, part (c)**

No we cannot use a deviance or log-likelihood ratio test to compare these two models because the link functions are different and they do not belong to the same family of distributions.

**Question 4, part (d)**

```r
model_d <- glm(dead_proportion ~ grp + hb, family = quasibinomial(link = "log"), data = data, weights =
summary(model_d)
```

```
##
## Call:
## glm(formula = dead_proportion ~ grp + hb, family = quasibinomial(link = "log"),
##     data = data, weights = data$N)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.7041  -1.1466  -0.0467   1.3622   2.9886
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.09939    0.21773   0.456 0.649848
## grpTreatment -1.89885    0.48810  -3.890 0.000272 ***
## hb           -0.08149    0.04914  -1.658 0.102980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 2.855775)
##
##     Null deviance: 509.43  on 57  degrees of freedom
## Residual deviance: 167.70  on 55  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 7
```

```
confint.default(model_d)
```

```
##                   2.5 %      97.5 %
## (Intercept)  -0.3273621  0.52614163
## grpTreatment -2.8555104 -0.94219932
## hb           -0.1778060  0.01483202
```

```
confint_a<-confint(model_a)
```

```
## Waiting for profiling to be done...
```

```
confint_a
```

```
##                  2.5 %     97.5 %
## (Intercept)   1.306244  2.7910677
## grpTreatment -3.634268 -1.7354767
## hb           -0.336825 -0.0446353
```

```
confint_b<-confint(model_b)
```

```
## Waiting for profiling to be done...
```

```
confint_b
```

```
##                   2.5 %      97.5 %
## (Intercept)  -0.1437809  0.32653251
## grpTreatment -2.4758065 -1.36361242
## hb           -0.1359477 -0.02896577
```

The model we should be using is the overdispersion model because it accounts for the extra variation present in the data, so it by using it, it makes sure that the standard errors and confidence intervals for each parameter are estimated and adjusted for additional variability.