Applied Logistic Regression, HW week5

homework 5
##Exercise One

For this exercise, you will use the Hyponatremia dataset. You can download the hyponatremia.dta Stata file, or you can also access the data through this CSV file.

The data for this homework exercise derive from an epidemiological study of hyponatremia (a life-threatening condition) among runners of the 2002 Boston Marathon. Hyponatremia is defined as an electrolyte disturbance in which the serum sodium concentration is lower than normal (<135 mmol/l).

The aim of the study was to determine whether a runner experienced hyponatremia and to identify the principal risk factors. Participants in the 2002 Boston Marathon completed a survey including demographic and anthropometric characteristics (Body Mass Index) one or two days before the race. After the race, runners provided a blood sample in order to measure their serum sodium concentration and completed a questionnaire detailing their urine output during the race. Pre-race and post-race weights were also recorded.

```
hypoNa <- read.csv("C:/Users/Karmen/Desktop/All Things Coursera/current/Applied Logistic
Regression/datasets/hyponatremia.csv")
str(hypoNa)
```

```
## 'data.frame':    488 obs. of  8 variables:
##  $ obs     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ na      : int  141 135 140 143 128 148 143 144 140 142 ...
##  $ nas135  : int  0 1 0 0 1 0 0 0 0 0 ...
##  $ female  : int  0 1 0 0 1 0 0 0 1 1 ...
##  $ urinat3p: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ wtdiff  : num  -2.87 0 -1.36 -1.82 -0.91 ...
##  $ bmi     : num  25.9 31.7 22.9 24.2 18.3 ...
##  $ runtime : int  230 400 216 201 223 275 309 197 253 260 ...
```

Complete the following:

a) Perform a logistic regression analysis using nas135 as dependent variable and female as the only independent variable. Interpret the coefficients of the model.
In `female` variable males are coded 0, and females are coded 1:

$$female = \begin{cases} 0 & \text{male} \\ 1 & \text{female} \end{cases}$$

```
logfit1 <- glm(nas135 ~ female, data = hypoNa, family = binomial(link = "logit"))
summary(logfit1)
```

```
##
## Call:
## glm(formula = nas135 ~ female, family = binomial(link = "logit"),
##     data = hypoNa)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7102  -0.7102  -0.4020  -0.4020   2.2608
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.4749     0.2082 -11.884  < 2e-16 ***
## female        1.2260     0.2795   4.386 1.16e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 371.60  on 487  degrees of freedom
## Residual deviance: 351.93  on 486  degrees of freedom
## AIC: 355.93
##
## Number of Fisher Scoring iterations: 5
```

```
#likelihood ratio test for significance of model with female as predictor compared to a n
aive model
(G1 <- logfit1$null.deviance - logfit1$deviance)
```

```
## [1] 19.66992
```

```
(df1 <- logfit1$df.null - logfit1$df.residual)
```

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

```
(pValue1 <- pchisq(G1, df = df1, lower.tail = FALSE))
```

```
## [1] 9.203901e-06
```

p-value of χ^2 test for likelihood ratio came out highly significant. We tested null hypothesis that model that predicts chances for experiencing hypnatremia based on gender of participant is equally significant as the naive model that has no predictors, against alternative hypothesis that model with gender as predictor is better than the model with no predictors. We rejected null hypothesis: model with `female` as predictor is better than the naive model.

```
(Wald1 <- summary(logfit1)$coefficients[2,1]/summary(logfit1)$coefficients[2,2])
```

```
## [1] 4.385547
```

```
(pValueW1 <- pnorm(Wald1, lower.tail = FALSE))
```

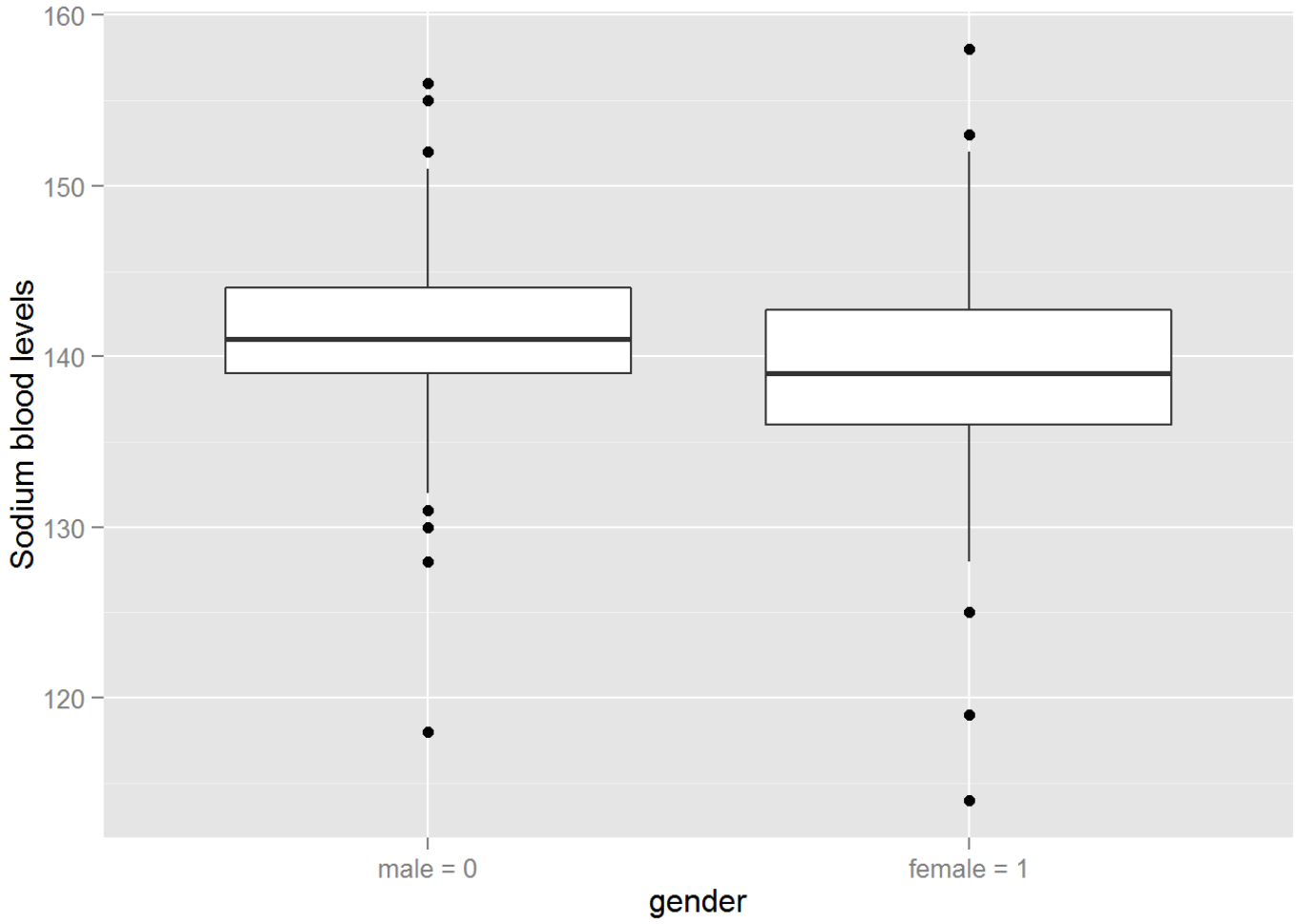
```
## [1] 5.784732e-06
```

Hypothesis test that tests whether slope of $\beta 1$ is different than zero yielded p-value that is significantly smaller than 0.05, which means slope for `female` is not zero.

```
(ORfem_male <- exp(logfit1$coefficients[2]))
```

```
##   female
## 3.407442
```

```
library(ggplot2)
ggplot(data = hypoNa, aes( x = factor(female,labels = c("male = 0", "female = 1")), y = n
a))   +
  geom_boxplot()  +
  xlab("gender")  +
  ylab("Sodium blood levels")
```



From the plot that compares sodium blood levels and gender we can see that women tend to have lower levels of sodium. Also, women are coded 1, and males are coded 0. Hence, if we calculate odds ratio from coefficients and their standard errors obtained from the output of logistic regression of `nas135` on `female` we can see that $\widehat{OR}(female, male) = 3.41$, which means that women have odds of developing hyponatremia 3.4 times higher than men. Value of Intercept denotes log odds for men to experience hyponatremia.
“The coefficient for female indicates the log odds ratio of hyponatremia of a female compared to a male. The constant indicates the log odds of hyponatremia for male. In this case, exp(1.225962)= 3.047, meaning that the odds of hyponatremia among females is 3.05
times that of males.”

b)Fit a model with runtime as the only independent variable. Interpret the coefficient for runtime.

```
logfit2 <- glm(nas135 ~ runtime, data = hypoNa, family = binomial(link = "logit"))
summary(logfit2)
```

```
##
## Call:
## glm(formula = nas135 ~ runtime, family = binomial(link = "logit"),
##     data = hypoNa)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1724  -0.5234  -0.4263  -0.3458   2.5182
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.592594   0.771282  -7.251 4.14e-13 ***
## runtime      0.015502   0.003091   5.015 5.29e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 360.90  on 476  degrees of freedom
## Residual deviance: 335.54  on 475  degrees of freedom
##   (11 observations deleted due to missingness)
## AIC: 339.54
##
## Number of Fisher Scoring iterations: 5
```

Slope for `runtime` indicates that log odds of hyponatremia increases by 0.015502 each minute it takes the marathon to be completed.

c)Calculate the Odds Ratio for the variable runtime and interpret it.

```
(OR <- exp(0.015502))
```

```
## [1] 1.015623
```

Odds ratio is ratio between odds of hypnatremia for the runner who took `x + 1` minutes to complete the marathon and the one who took `x` minutes to complete the marathon.
“The ratio of the odds of hyponatremia of a runner who is 1 minute slower than another runner is 1.0156. This indicates that the odds of hyponatremia increases by 1.56% for each additional minute it takes to complete the marathon. This ratio is constant no matter what values of runtime are compared, provided that the difference in time is 1 minute (eg. 201 minutes vs 200 minutes, 251 minutes vs 250 minutes, etc.)”

d)Interpret the coefficient for the constant in the model with runtime as the only independent variable. Does it make sense? If not, what can you do to obtain a coefficient for the constant which is easily interpreted?
Intercept represents log odds for hyponatremia for runner that took 0 minutes to complete the marathon. Since that is nonsensical, intercept only reflects slope and height of the line.
If we center variable `runtime` so to subtract mean value of `runtime` from each observation, then

intercept would mark log odds of hyponatremia for a runner that completes marathon in average runtime.
“This coefficient gives the log odds of hyponatremia of a runner who takes 0 minutes to complete the marathon. It does not make any sense! If runtime were centered around the mean (i.e. if a new variable called runtime2 were created by subtracting the mean value of runtime to all observations) then the coefficient would have indicated the log odds of hyponatremia of a runner who completes the marathon in the average time.”

e)Calculate the Odds Ratio of hyponatremia of a runner who takes 2 hours more than another runner, and the corresponding 95% Confidence Interval

```
#logit for runner that ran 120 minutes longer than the other runner
(lg_2hrs <-logfit2$coefficients[1] + logfit2$coefficients[2] * 121)
```

```
## (Intercept)
##    -3.71686
```

```
#the other runner who ran 120 minutes less than the previous runner
(lg_1 <- logfit2$coefficients[1] + logfit2$coefficients[2] * 1)
```

```
## (Intercept)
##   -5.577092
```

```
#log odds ratio of these two runners
(G <- lg_2hrs - lg_1)
```

```
## (Intercept)
##    1.860231
```

```
#odds ratio for a runner that runs 120 minutes longer than the other runner
(OR_2hrs <- exp(G))
```

```
## (Intercept)
##    6.425222
```

```
#CI for a log odds of developing hyponatremia for a runner that runs 120 minutes longer
(ci <- G + c(-1,1)*qnorm(0.975) * summary(logfit2)$coefficients[2,2]*120)
```

```
## [1] 1.133273 2.587190
```

```
#CI for odds(odds ratio) of developing hyponatremia for runner that runs 120 minutes long
er than the other runner
exp(ci)
```

```
## [1]  3.105804 13.292365
```

Same thing we can acomplish by simply multiplying β_1 with 120 (minutes) and then exponentiating that amount: this would give us odds ratio for developing hyponatriemia in runner who ran 120 minutes longer then the other runner. To build confidence interval we need to simply build a 95% confidence interval around $\beta_1 * 120$. We do that by multiplying standard error by 120:

$$CI = \beta_1 \pm z_{(1-\alpha)/2} * SE_{\beta_1} * 120$$

```
(beta_1 <- logfit2$coefficients[2])
```

```
##    runtime
## 0.01550193
```

```
(x <-  beta_1 * 120)
```

```
##  runtime
## 1.860231
```

```
exp(beta_1 * 120)
```

```
##  runtime
## 6.425222
```

```
(CI_2hrs <- x + c(-1,1)*qnorm(0.975)*summary(logfit2)$coefficients[2,2]*120)
```

```
## [1] 1.133273 2.587190
```

```
(OR_2hrs_CI <- exp(CI_2hrs))
```

```
## [1]  3.105804 13.292365
```

The odds of hyponatremia for a runner who takes 2 hours more to complete the marathon are about 3 to 13 % higher than odds for the runner who is faster, or on average about 6% higher.

On the other hand we can extract default confidence interval from a model with `confint.default(model)` and multiply it by 120:

```
confint.default(logfit2)
```

```
##                   2.5 %      97.5 %
## (Intercept) -7.104278112 -4.08090898
## runtime      0.009443939  0.02155991
```

```
exp(confint.default(logfit2)[2,] * 120)
```

```
##     2.5 %    97.5 %
##  3.105804 13.292365
```

f)Fit a model with female and runtime as independent variables. Interpret both coefficients.

```
logfit3 <- glm(nas135 ~ female + runtime, data = hypoNa, family = binomial(link = "logi
t"))
summary(logfit3)
```

```
##
## Call:
## glm(formula = nas135 ~ female + runtime, family = binomial(link = "logit"),
##     data = hypoNa)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2157  -0.5714  -0.3668  -0.2899   2.6427
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.721056   0.823279  -6.949 3.68e-12 ***
## female       0.963836   0.291048   3.312 0.000928 ***
## runtime      0.014214   0.003295   4.314 1.60e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 360.90  on 476  degrees of freedom
## Residual deviance: 324.48  on 474  degrees of freedom
##   (11 observations deleted due to missingness)
## AIC: 330.48
##
## Number of Fisher Scoring iterations: 5
```

Slopes for the coefficients mean the same they meant when they were the only predictor, but their value is changed as a result of presence of another variable in the model. In that sense, slope for female is log odds of developing hyponatremia in females compared to males, all else held constant. Slope for

runtime is change in log odds of developing hyponatremia for one minute increase in runtime, all else held constant.
Compare the coefficients for female in the model with female as the only independent variable with that in the model that contains female and runtime. What is the percentage change in the coefficient of female?
“The coefficient for female is the log odds of hyponatremia of a female compared to a male who completes the marathon in the same time. The coefficient for runtime is the log odds of hyponatremia for an additional minute that it takes to complete the marathon, independent from the fact that the runner is a male or a female.”

g)Compare the coefficients for female in the model with female as the only independent variable with that in the model that contains female and runtime. What is the percentage change in the coefficient of female?

```
coefUNI <- logfit1$coefficients["female"]
coefMLR <- logfit3$coefficients["female"]

(percentage_change <- ((coefUNI - coefMLR)/coefUNI) * 100)
```

```
##  female
## 21.3812
```

There is about 21.4 % change in the coefficients.

h)Calculate the Odds Ratio of hyponatremia for a female compared to a male who completes the marathon in the same time.

```
logfit3$coefficients
```

```
## (Intercept)      female     runtime
## -5.72105597  0.96383641  0.01421362
```

```
beta_1 <- logfit3$coefficients[2]
(OR <- exp(beta_1))
```

```
##   female
## 2.621735
```

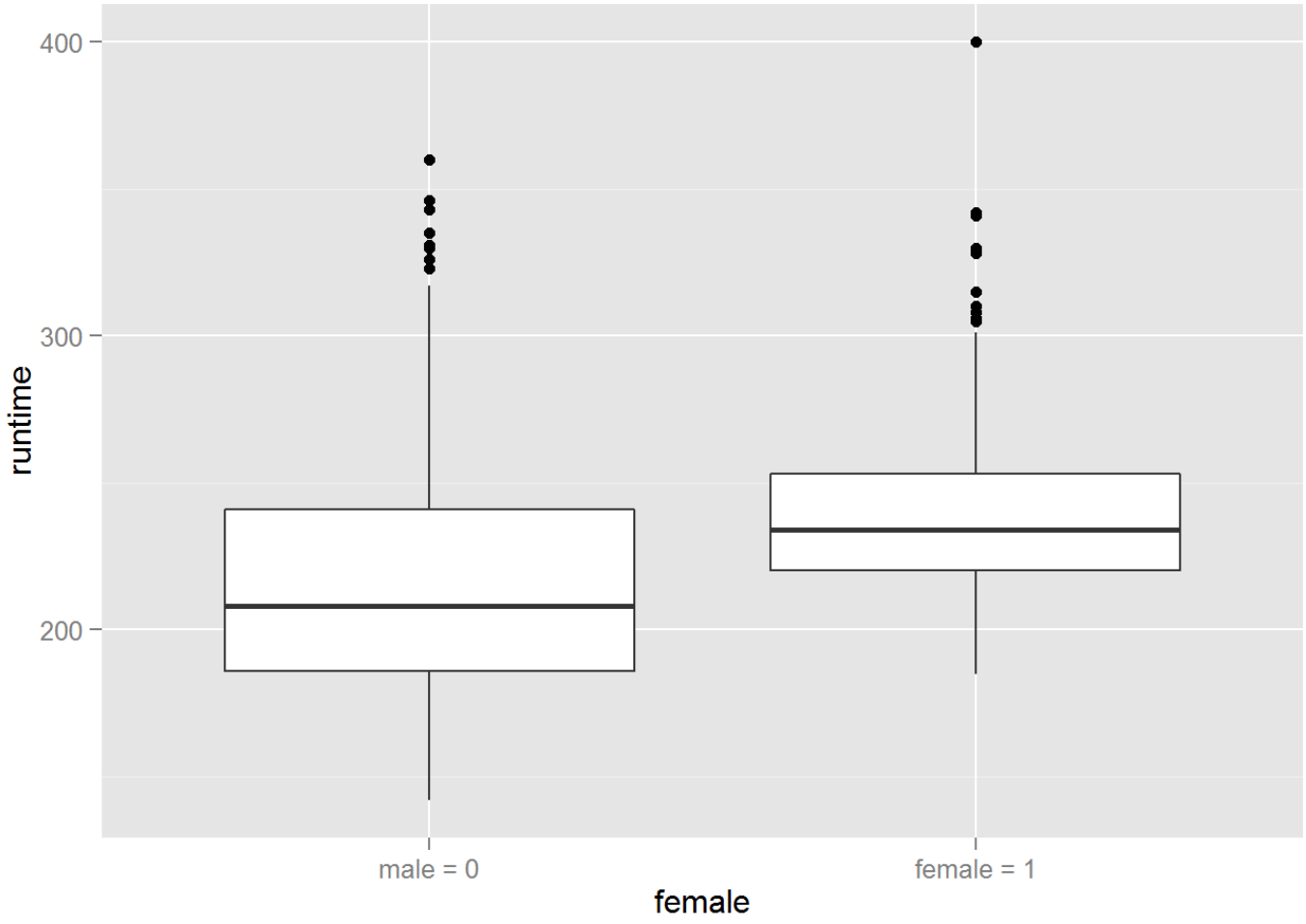
Woman that completes marathon in same time as her male colleague has about 2.62 times greater odds to develop hyponatremia then that man.

i)What type of association do you expect between the variables female and runtime? Answer this question before looking at the data, only on the basis of the observed change in the coefficient for female when runtime is entered into the model. Then make a box-plot of runtime by female.
“We expect **a positive association** between female and runtime: on average females will be slower

than males. This can be deduced because the coefficient for female decreases when runtime is entered into the model and because runtime has a positive association itself with nas135. Part of the effect of female on nas135 in the univariable model is confounded by the positive association between female and runtime. The box-plot makes clear this association."

```
ggplot(data = hypoNa, aes(x = factor(female, labels = c("male = 0", "female = 1")), y = r
untime))  +
  geom_boxplot()  +
  xlab("female")  +
  ylab("runtime")
```

```
## Warning in loop_apply(n, do.ply): Removed 11 rows containing non-finite
## values (stat_boxplot).
```



j)Assess whether there is an interaction between female and runtime.

```
logfit4 <- glm(nas135 ~ female * runtime, data = hypoNa, family = binomial(link = "logi
t"))
summary(logfit4)
```

```
##
## Call:
## glm(formula = nas135 ~ female * runtime, family = binomial(link = "logit"),
##     data = hypoNa)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1505  -0.5874  -0.3611  -0.2798   2.6786
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.006665   1.067610  -5.626 1.84e-08 ***
## female          1.664387   1.666037   0.999 0.317790
## runtime         0.015392   0.004299   3.581 0.000343 ***
## female:runtime -0.002845   0.006657  -0.427 0.669123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 360.9  on 476  degrees of freedom
## Residual deviance: 324.3  on 473  degrees of freedom
##   (11 observations deleted due to missingness)
## AIC: 332.3
##
## Number of Fisher Scoring iterations: 5
```

```
(logLik_logfit3 <- logLik(logfit3))
```

```
## 'log Lik.' -162.2398 (df=3)
```

```
(logLik_logfit4 <- logLik(logfit4))
```

```
## 'log Lik.' -162.1488 (df=4)
```

```
(G <- -2*(logLik_logfit3 -  logLik_logfit4))
```

```
## 'log Lik.' 0.182022 (df=3)
```

```
(df_Interaction <- logfit3$df.residual - logfit4$df.residual)
```

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

```
(pValueInteraction <- pchisq(as.numeric(as.character(G)), df = df_Interaction, lower.tail
= FALSE))
```

```
## [1] 0.6696413
```

p-value from the χ^2 test came out non-significant, 0.669. Hence slope for the interaction term β_3 is not significantly different from 0.

** k)Add to the model that contains female and runtime a dichotomous variable wgain which takes the value of 0 if wtidff = 0, and the value of 1 if wtidff > 0. Test for interaction between female and wgain.**

```
hypoNa$wgain <- ifelse(hypoNa$wtdiff <= 0, 0, 1)
logfit5 <- glm(nas135 ~ female + runtime + wgain, data = hypoNa, family = binomial(link =
"logit"))
logfit6 <- glm(nas135 ~ runtime + female*wgain, data = hypoNa, family = binomial(link =
"logit"))

logLik_logfit5 <- logLik(logfit5)
logLik_logfit6 <- logLik(logfit6)
G_56 <- -2*(logLik_logfit5 - logLik_logfit6)

df_56 <- logfit5$df.residual - logfit6$df.residual
(pValue_56 <- pchisq(G_56, df = df_56, lower.tail = FALSE))
```

```
## 'log Lik.' 0.06462293 (df=4)
```

p-value from χ^2 test is 0.06462, which is less than 10%. Therefore, we can conclude that `wgain` is effect modifier, rather than just a simple confounder.
“The coefficient for the interaction term is significant at the 10% level (p=0.069>0.1)”

l)On the basis of the model with the interaction term, calculate the Odds Ratios of hyponatremia for males who gain weight as compared to those who don’t. Repeat this exercise for a female. Interpret your findings.
Formula for logit for this model is

$$g = \beta_0 + \beta_1 * female + \beta_2 * runtime + \beta_3 * wgain + \beta_4 female : wgain$$

Logit for male that gained weight:

$$g(male, gained) = \beta_0 + \beta_1 * 0 + \beta_2 * runtime + \beta_3 * 1 + \beta_4 * 0 * 1$$

$$g(male, gained) = \beta_0 + \beta_2 * runtime + \beta_3 * 1$$

Logit for male that didn’t gain weight:

$$g(male, didn't gain) = \beta_0 + \beta_1 * 0 + \beta_2 * runtime + \beta_3 * 0 + \beta_4 * 0 * 0$$

$$g(male, didn't gain) = \beta_0 + \beta_2 * runtime$$

Then log odds ratio is

$$G = g(male, gained) - g(male, didn't gain)$$

$$G = \beta_0 + \beta_2 * runtime + \beta_3 * 1 - \backslash[\beta_0 + \beta_2 * runtime$$

\]

$$G = \beta_0 + \beta_2 * runtime + \beta_3 - \beta_0 - \beta_2 * runtime$$

$$G = \beta_3$$

Odds ratio is:

$$\widehat{OR}(malegained, maledidn't gain) = e^{\beta_3}$$

Logit for female that gained weight:

$$g(female, gained) = \beta_0 + \beta_1 * 1 + \beta_2 * runtime + \beta_3 * 1 + \beta_4 * 1 * 1$$

Logit for female that didn't gainOR:

$$g(female, didn't gain) = \beta_0 + \beta_1 * 1 + \beta_2 * runtime + \beta_3 * 0 + \beta_4 * 1 * 0$$

$$g(female, didn't gain) = \beta_0 + \beta_1 * 1 + \beta_2 * runtime$$

Then log odds ratio is:

$$G = g(female, gained) - g(female, didn't gain)$$

$$G = \beta_0 + \beta_1 * 1 + \beta_2 * runtime + \beta_3 * 1 + \beta_4 * 1 * 1 - \beta_0 + \beta_1 * 1 + \beta_2 * runtime$$

\]

$$G = \beta_3 + \beta_4$$

Odds ratio is:

$$\widehat{OR}(femalegained, femaledidn't gain) = e^{\beta_3 + \beta_4}$$

```
(coefs <- summary(logfit6)$coef)
```

```
##                 Estimate  Std. Error   z value     Pr(>|z|)
## (Intercept)  -7.53349526 1.069754861 -7.042263 1.891421e-12
## runtime       0.01687962 0.003889585  4.339696 1.426803e-05
## female        1.50083382 0.522473405  2.872555 4.071666e-03
## wgain         2.40100017 0.511935491  4.690044 2.731458e-06
## female:wgain -1.20185577 0.660934758 -1.818418 6.900024e-02
```

```
(beta_2 <- coefs[3,1])
```

```
## [1] 1.500834
```

```
(beta_3 <- coefs[4,1])
```

```
## [1] 2.401
```

```
(beta_4 <- coefs[5,1])
```

```
## [1] -1.201856
```

```
#males who gain weight compared to those who don't
(OR_mgain <- exp(beta_3))
```

```
## [1] 11.03421
```

```
#females who gain weight compared to those who don't
(OR_fgain <- exp(beta_3 + beta_4))
```

```
## [1] 3.317277
```

Men that gained weight have about 11 times greater odds of developing hyponatremia than those who didn't gain weight, for the same runtime. Females that gained weight, on the other hand have about 3.3 times greater odds of developing hyponatremia compared to the females that didn't gain weight and have the same runtime.

m)Compare using the Likelihood Ratio test the model with female and runtime with a model with female, runtime, wgain, urinat3p and bmi. (Hint: the 2 models must be fitted on the same set of observations. Be aware of missing values in some of these variables). How many degrees of freedom does the test statistic have?

creating a subset of complete cases(no missing entries) for variables we need in the model:

```
full_indices <- complete.cases(hypoNa[, c("female", "runtime", "wgain", "urinat3p", "bm
i")])
log_full_subset <- hypoNa[full_indices, c("nas135", "female", "runtime", "wgain", "urinat
3p", "bmi")]
```

MODEL WITH FEMALE AND RUNTIME: we need to refit this model because in the next model we are working with much reduced dataset, and to compare model using likelihood ratio test the two models need to be fit on the same dataset

```
#MODEL WITH FEMALE AND RUNTIME:
logfit3_red <- glm(nas135 ~ female + runtime, data = log_full_subset, family = binomial(l
ink = "logit"))
summary(logfit3_red)
```

```
##
## Call:
## glm(formula = nas135 ~ female + runtime, family = binomial(link = "logit"),
##     data = log_full_subset)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2198  -0.5626  -0.3677  -0.2880   2.6724
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.959965   0.897346  -6.642 3.10e-11 ***
## female       0.873966   0.305045   2.865  0.00417 **
## runtime      0.015205   0.003589   4.236 2.27e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 328.17  on 441  degrees of freedom
## Residual deviance: 296.48  on 439  degrees of freedom
## AIC: 302.48
##
## Number of Fisher Scoring iterations: 5
```

```
#MODEL WITH FEMALE, RUNTIME, WGAIN, URINAT3P AND BMI
logfit7 <- glm(nas135 ~ female + runtime + wgain + urinat3p + bmi, data = log_full_subse
t, family = binomial(link = "logit"))
summary(logfit7)
```

```
##
## Call:
## glm(formula = nas135 ~ female + runtime + wgain + urinat3p +
##     bmi, family = binomial(link = "logit"), data = log_full_subset)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4099  -0.4711  -0.2969  -0.2000   2.8722
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.565610   1.599794  -4.104 4.06e-05 ***
## female       0.759657   0.415521   1.828  0.06752 .
## runtime      0.014701   0.004839   3.038  0.00238 **
## wgain        1.735328   0.330983   5.243 1.58e-07 ***
## urinat3p     0.815514   0.551410   1.479  0.13915
## bmi         -0.004152   0.074235  -0.056  0.95540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 328.17  on 441  degrees of freedom
## Residual deviance: 263.23  on 436  degrees of freedom
## AIC: 275.23
##
## Number of Fisher Scoring iterations: 6
```

```
#likelihood ratio test:
(G_2 <- logfit3_red$deviance - logfit7$deviance)
```

```
## [1] 33.24758
```

```
#degrees of freedom for the likelihood ratio test:
(df_G_2 <- logfit3_red$df.residual - logfit7$df.residual)
```

```
## [1] 3
```

```
#finding p-value
(pVal <- pchisq(G_2, df = df_G_2, lower.tail = FALSE))
```

```
## [1] 2.85574e-07
```

χ^2 test evaluated at 3 degrees of freedom yielded p-value that is highly significant. Therefore model with 5 predictors is better than model with 2 predictors.