Homework Solutions
Applied Logistic Regression

WEEK 4

Exercise 1 (continued):

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

To calculate the odds ratio, exponentiate the logit:

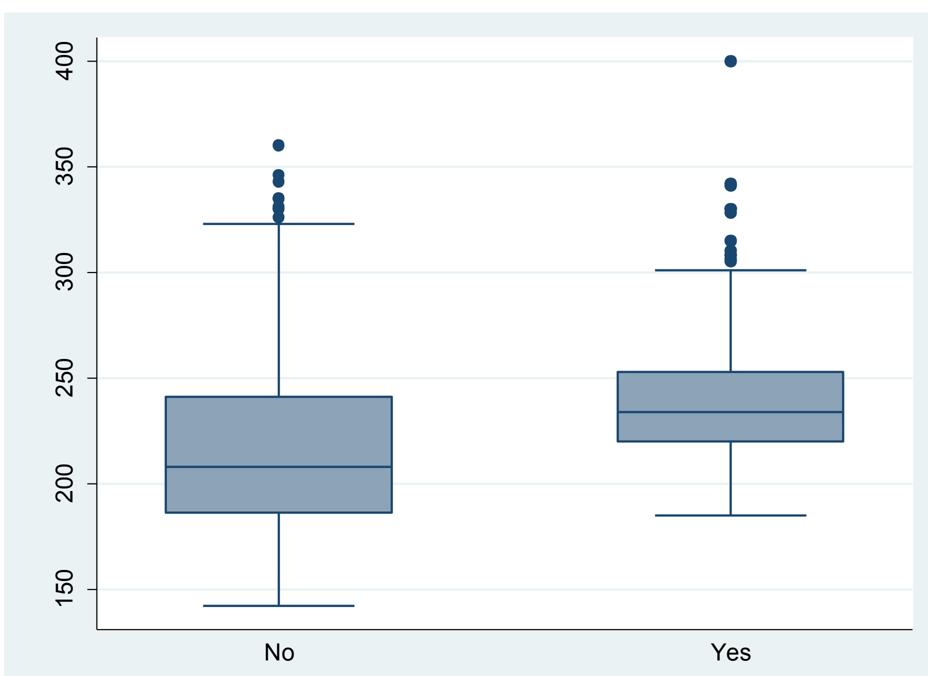
```
. di exp(0.9638)
2.6216398
```

The odds ratio is 2.62.

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.

```
. graph box runtime, over(female)
```



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

For this part of the analysis, you must first generate an interaction term:

```
. gen femXrun=female*runtime
(11 missing values generated)
```

Next, fit the regression with the interaction term included:

```
. logit nas135 female runtime femXrun, nolog

Logistic regression                               Number of obs   =        477
                                                  LR chi2(3)      =      36.60
                                                  Prob > chi2     =     0.0000
Log likelihood = -162.14884                       Pseudo R2       =     0.1014
------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.664386   1.666064     1.00   0.318    -1.601039    4.929811
     runtime |    .015392   .0042988     3.58   0.000     .0069666    .0238175
     femXrun |  -.0028449   .0066572    -0.43   0.669    -.0158929     .010203
       _cons |  -6.006664   1.067652    -5.63   0.000    -8.099224   -3.914104
```

NOTE: An alternative command for Stata 11 is:

```
. logit nas135 i.female##c.runtime, nolog

Logistic regression                               Number of obs   =        477
                                                  LR chi2(3)      =      36.60
                                                  Prob > chi2     =     0.0000
Log likelihood = -162.14884                       Pseudo R2       =     0.1014

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |   1.664386   1.666064     1.00   0.318    -1.601039    4.929811
     runtime |    .015392   .0042988     3.58   0.000     .0069666    .0238175
             |
     female#|
   c.runtime |
          1  |  -.0028449   .0066572    -0.43   0.669    -.0158929     .010203
             |
       _cons |  -6.006664   1.067652    -5.63   0.000    -8.099224   -3.914104
------------------------------------------------------------------------------
```

The interaction term between the 2 variables is far from significant (p=0.669>>0.05). There is no interaction between these 2 variables.

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**.

First, generate a new variable (wgain), making sure to generate missing variables for wtgain in the event that observations of wtdiff that are missing (Stata recognizes missing variables as having a value of positive infinity). Run a regression with wgain included.

```
. gen wgain=wtdiff>0

. replace wgain=. if wtdiff==.
(33 real changes made, 33 to missing)

. logit nas135 female runtime wgain, nolog

Logistic regression                               Number of obs   =        449
                                                  LR chi2(3)      =      67.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -135.31522                       Pseudo R2       =     0.1990

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .7763378   .3166739     2.45   0.014     .1556684    1.397007
     runtime |   .0168388   .0038002     4.43   0.000     .0093905     .024287
       wgain |   1.729974   .3234668     5.35   0.000     1.095991    2.363957
       _cons |  -7.105477   1.001043    -7.10   0.000    -9.067485   -5.143469
------------------------------------------------------------------------------
```

Generate an interaction term between gender and weight gain. Run a regression with the interaction term included as well:

```
------------------------------------------------------------------------------

. gen femXgain=female*wgain
(33 missing values generated)

. logit nas135 female runtime wgain femXgain, nolog

Logistic regression                               Number of obs   =        449
                                                  LR chi2(4)      =      70.64
                                                  Prob > chi2     =     0.0000
Log likelihood = -133.60793                       Pseudo R2       =     0.2091

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.500834     .52248     2.87   0.004     .4767919    2.524876
     runtime |   .0168796   .0038896     4.34   0.000     .0092561    .0245031
       wgain |      2.401   .5119424     4.69   0.000     1.397612    3.404389
    femXgain |  -1.201856   .6609401    -1.82   0.069    -2.497275     .093563
       _cons |  -7.533495   1.069764    -7.04   0.000    -9.630193   -5.436797
```

The coefficient for the interaction term is significant at the 10% level (p=0.069>0.1)

I. 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.

Based on this model, the logit for females is

$$\beta_0 + \beta_1(female) + \beta_2(runtime) + \beta 3(wgain) + \beta_4(fem \times gain)$$

So, for females, the log odds ratio (= logit difference) comparing those with weight gain vs those without is

$$\left[\beta_0 + \beta_1(1) + \beta_2(runtime) + \beta_3(1) + \beta_4(1 \times 1)\right] - \left[\beta_0 + \beta_1(1) + \beta_2(runtime) + \beta_3(0) + \beta_4(1 \times 0)\right]$$

$$= \beta_3 + \beta_4 = 2.401 - 1.202$$

and the odds ratio = $e^{2.401-1.202} = 3.317$

For males, the log odds ratio (= logit difference) comparing those with weight gain vs those without is:

$$\left[\beta_0 + \beta_1(0) + \beta_2(runtime) + \beta_3(1) + \beta_4(0 \times 1)\right] - \left[\beta_0 + \beta_1(0) + \beta_2(runtime) + \beta_3(0) + \beta_4(0 \times 0)\right]$$

$$= \beta_3 = 2.401$$

and the odds ratio = $e^{2.401} = 11.034$

```
. di exp(2.401)
11.034205

. di exp(2.401-1.202)
3.3167985
```

A male who experiences weight gain during a marathon has an odds of hyponatremia about 11 times higher than that of a male who does not gain weight. On the other hand, a female who experiences weight gain during a marathon has an odds of hyponatremia about 3 times higher than that of a female who does not gain weight.

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?

First, generate a subpopulation (nomiss) for all of the observations without missing variables (Note:"!=." is code for "does not equal to missing")

```
. gen nomiss=0

. replace nomiss=1 if female!=. & urin!=. &bmi!=. &wgain!=. &runtime!=.
(442 real changes made)
```

Run the full model, and store the estimates for the model under the name "A" using the command "est store A"

```
. logit nas135 female runtime wgain bmi urina, nolog

Logistic regression                               Number of obs   =        442
                                                  LR chi2(5)      =      64.93
                                                  Prob > chi2     =     0.0000
Log likelihood = -131.61627                       Pseudo R2       =     0.1979

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .7596571   .4155214     1.83   0.068    -.0547497    1.574064
     runtime |   .0147009   .0048388     3.04   0.002     .0052171    .0241848
       wgain |   1.735328    .330983     5.24   0.000     1.086613    2.384043
         bmi |  -.0041517   .0742347    -0.06   0.955    -.1496491    .1413456
    urinat3p |   .8155137   .5514101     1.48   0.139    -.2652302    1.896258
       _cons |   -6.56561   1.599794    -4.10   0.000    -9.701149    -3.43007
------------------------------------------------------------------------------

. est store A
```

Run a second regression with only female and runtime as independent variables, making sure to exclude all missing variables by limiting the analysis to the subpopulation ("nomiss==1"). Store the estimates of the model under the name "B" through the command "est store B"

```
. logit nas135 female runtime if nomiss==1, nolog

Logistic regression                               Number of obs   =        442
                                                  LR chi2(2)      =      31.69
                                                  Prob > chi2     =     0.0000
Log likelihood = -148.24005                       Pseudo R2       =     0.0966

------------------------------------------------------------------------------
      nas135 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .8739657   .3050463     2.87   0.004      .276086    1.471845
     runtime |   .0152055   .0035895     4.24   0.000     .0081702    .0222408
       _cons |  -5.959965   .8973516    -6.64   0.000    -7.718742   -4.201188
------------------------------------------------------------------------------

. est store B
```

Run a likelihood ratio test comparing the estimates of the full model (A) to those of the reduced model (B) through the command "lrtest B A"

```
. lrtest B A

Likelihood-ratio test                                 LR chi2(3)  =     33.25
(Assumption: B nested in A)                           Prob > chi2 =    0.0000
```

The LR test is highly significant. The test uses 3 degrees of freedom, which is the difference in the number of covariates (5-2=3). The model with 5 covariates is better than the one with 2 covariates.