# THE ANALYSIS OF HIERARCHICAL AND OTHER DEPENDENT DATA

**Solutions to Computer Practicals**

**LSHTM MSc in Medical Statistics**

# Practical 0: Revision of linear regression modelling
# Solutions

## Questions

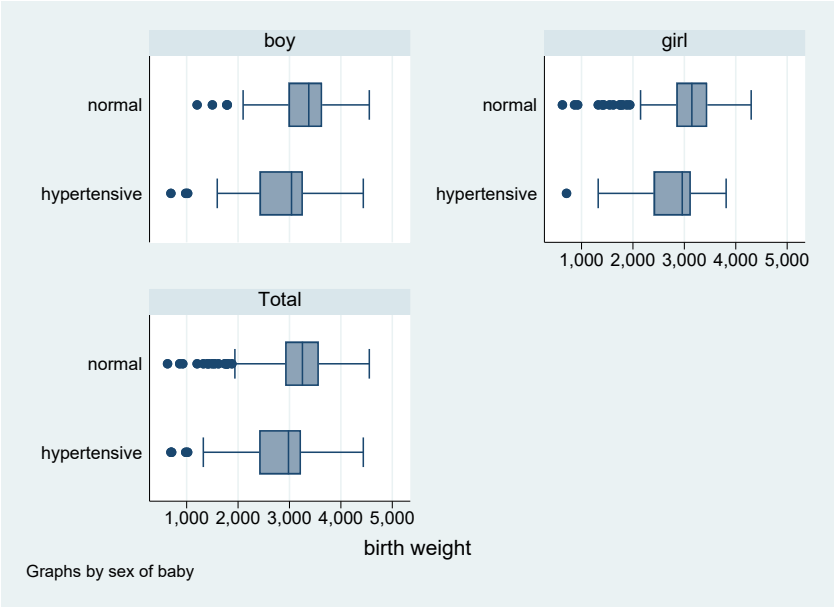
1. The dataset holds records on 500 babies. The variable `gestwk` has 10 missing values; all variable ranges seem plausible.

2. After labelling `hyp` and `sex` we cross-tabulate them:

```
. tab hyp sex
             |      sex of baby
   hypertens |       boy       girl |     Total
-------------+----------------------+----------
      normal |       221        207 |       428
hypertensive |        43         29 |        72
-------------+----------------------+----------
       Total |       264        236 |       500
```

3. Both boys and girls born to hypertensive mothers have lower birth weight (by 500-600$g$) than those born to mothers with normal blood pressure (see Figure 1).

Figure 1: Box and whiskers plot of birthweight by hypertension and sex




The corresponding sample statistics are:

```
. tabstat bw if sex==1,by(hyp) s(count mean sd)
 hyp |         N      mean        sd
-------------+------------------------------
      normal |       221  3310.747  554.8759
hypertensive |        43  2814.395  830.8337
-------------+------------------------------
       Total |       264  3229.902  633.6428
------------------------------------------

. tabstat bw if sex==2,by(hyp) s(count mean sd)
        hyp |         N      mean        sd
-------------+------------------------------
      normal |       207  3079.498  597.2287
hypertensive |        29  2699.724  736.4475
-------------+------------------------------
       Total |       236  3032.831   626.816
------------------------------------------
```

4. On the basis of the t-test and F-test statistics, we reject the hypothesis that the mean birth weight of babies is the same for hypertensive mothers and mothers with normal blood pressure.(Note however that the group specific SDs are quite different.)

```
. ttest bw, by(hyp)
Two-sample t test with equal variances
------------------------------------------------------------------------------
```

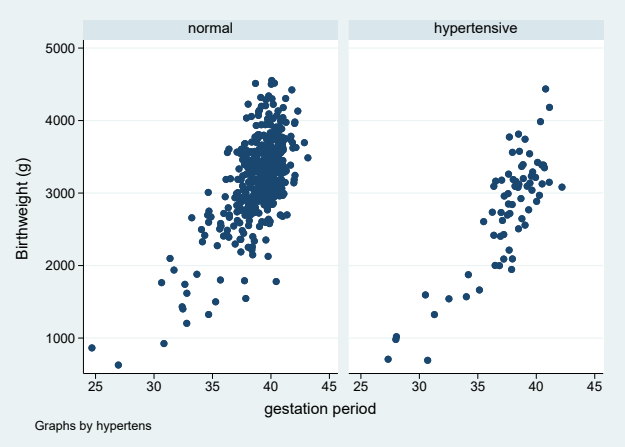
```
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
normativ |     428    3198.904     28.3541    586.5941    3143.173    3254.635
hyperten |      72    2768.208    93.20165    790.8422     2582.37    2954.047
---------+--------------------------------------------------------------------
combined |     500    3136.884     28.5077    637.4515    3080.874    3192.894
---------+--------------------------------------------------------------------
    diff |            430.6959    78.95458                275.5707     585.821
------------------------------------------------------------------------------
diff = mean(normativ) - mean(hyperten)                        t =   5.4550
Ho: diff = 0                                     degrees of freedom =      498
Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000

. oneway bweight hyp
Analysis of Variance
Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      11432670.3      1   11432670.3     29.76     0.0000
 Within groups       191333183    498   384203.179
------------------------------------------------------------------------
    Total            202765853    499   406344.395

Bartlett's test for equal variances:  chi2(1) =  12.4254  Prob>chi2 = 0.000
```

5. There seems to be a slightly stronger relation between birth weight and gestational age in babies born to hypertensive mothers. However the numbers are small (see Figure 2).

Figure 2: birthweight versus gestational period by hypertension




6. We first centre gestational period around its mean and generate the variable `c_gest`.

   The unadjusted regression coefficient for hypertension is large and highly significant (-431$g$ is the expected reduction in mean birthweight). However this is confounded by the effect of gestational period, as the adjusted coefficient reduced to -144$g$ in the model that also includes gestational period.

There is no evidence of effect modification between these two variables (Wald test statistic for interaction=1.55). Maternal hypertension appears to be a minor confounder for the effect of gestational period as the regression coefficient of the latter decreases from $197g$ to $192g$ when controlled for hypertension.

```
. regress bweight c_gest
------------------------------------------------------------------------------
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      c_gest |   196.9726   8.788133    22.41   0.000     179.7054    214.2399
       _cons |   3138.006   20.31645   154.46   0.000     3098.088    3177.925
------------------------------------------------------------------------------

regress bweight hyp
------------------------------------------------------------------------------
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hyp |  -430.6959   78.95458    -5.45   0.000     -585.821   -275.5707
       _cons |   3198.904   29.96116   106.77   0.000     3140.038     3257.77
------------------------------------------------------------------------------

. regress bweight hyp c_gest
------------------------------------------------------------------------------
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hyp |  -143.6749   58.81996    -2.44   0.015    -259.2472    -28.1027
      c_gest |   192.2384   8.955992    21.46   0.000     174.6412    209.8355
       _cons |   3158.824   21.93717   143.99   0.000     3115.721    3201.928
------------------------------------------------------------------------------

. regress bweight i.hyp##c.c_gest

      Source |       SS       df       MS              Number of obs =     490
-------------+------------------------------           F(  3,   486) =  172.44
       Model |   103277702     3  34425900.8           Prob > F      =  0.0000
    Residual |  97024840.5   486   199639.59           R-squared     =  0.5156
-------------+------------------------------           Adj R-squared =  0.5126
       Total |   200302543   489  409616.652           Root MSE      =  446.81

------------------------------------------------------------------------------
     bweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hyp |
hypertensive |  -117.3668   61.14039    -1.92   0.055    -237.4989    2.765336
      c_gest |   183.9105   10.43394    17.63   0.000     163.4093    204.4117
             |
 hyp#c.c_gest |
hypertensive |   31.38511   20.25549     1.55   0.122    -8.414033    71.18424
             |
       _cons |   3160.539   21.93363   144.10   0.000     3117.443    3203.636
------------------------------------------------------------------------------
```

7. The residuals seem to support the normality assumptions of the model (see Figure 3).

Figure 3: Residuals from the linear regression model of birth weight on gestational age, maternal hypertension, gender and maternal age