Answer 8.81

All missing data are coded as 9. We will exclude the missing values with the following commands:

```
. replace painmx_2=. if painmx_2==9
(1 real change made, 1 to missing)

. replace painmx_4=. if painmx_4==9
(5 real changes made, 5 to missing)
```

To calculate the difference in degree of pain, we create a new variable to represent the difference in degree of pain while receiving drug compared to placebo:

```
. gen diffmx = (painmx_4 - painmx_2)*(-1)^(drg_ord)
(5 missing values generated)
```

a) We assess the normality of this variable to determine whether we can apply the central limit theorem:

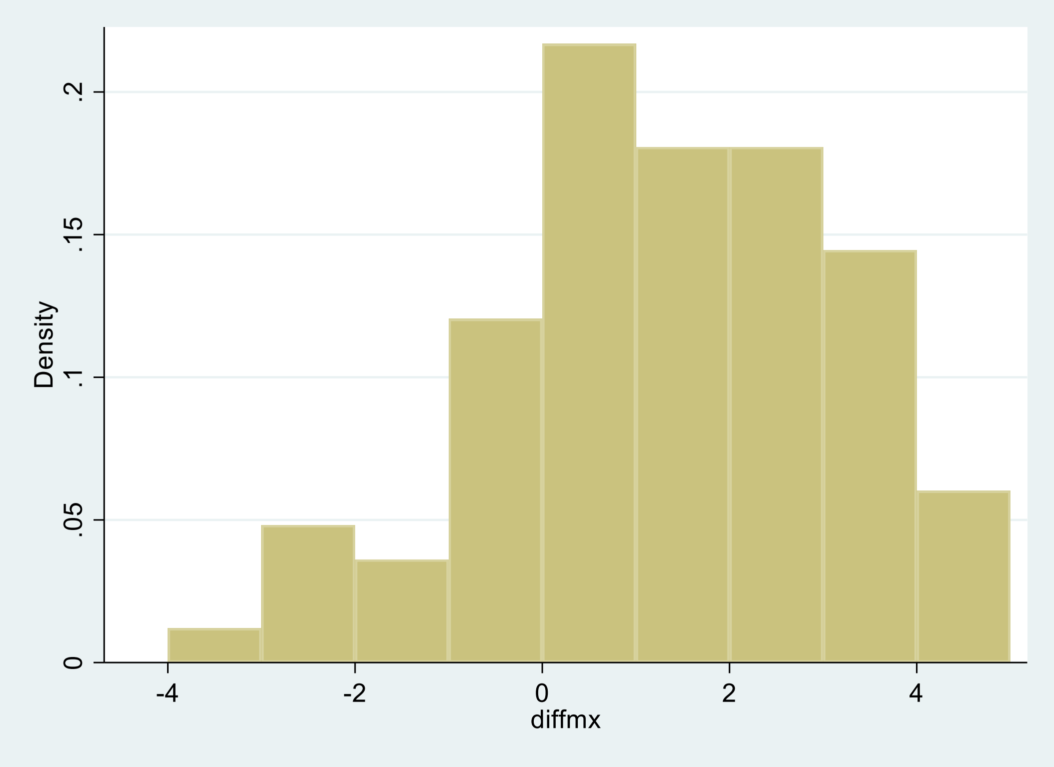


Figure 1. The histogram of the difference in degree of pain.

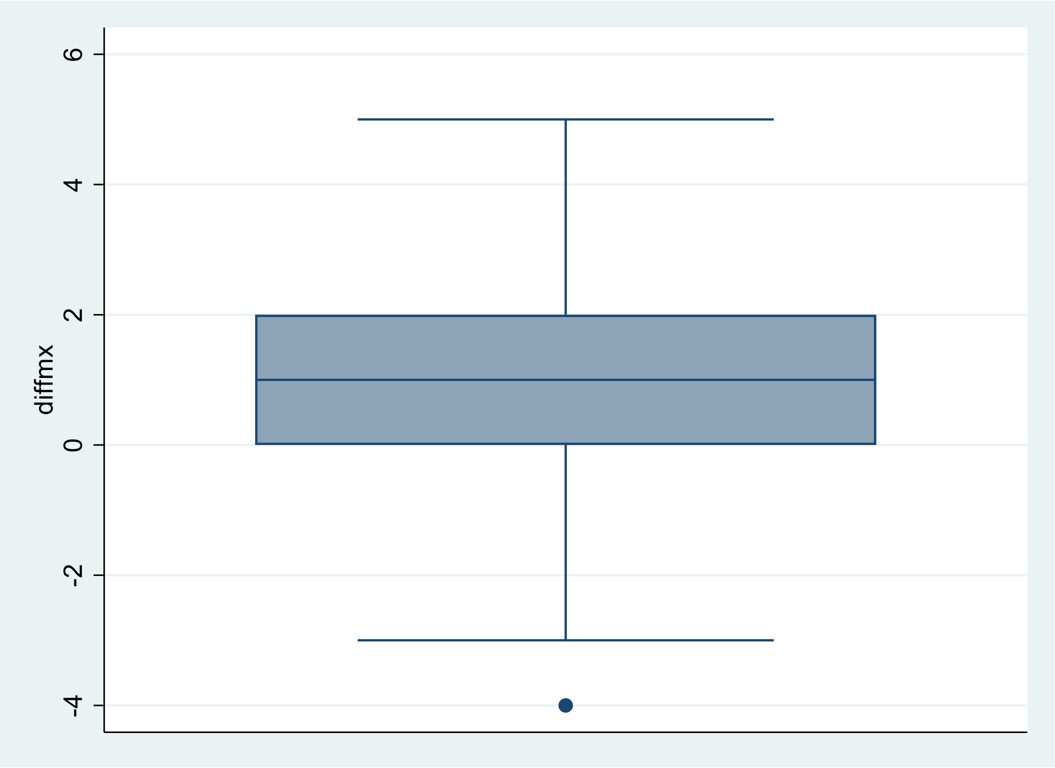


Figure 2. The boxplot of the difference in degree of pain.

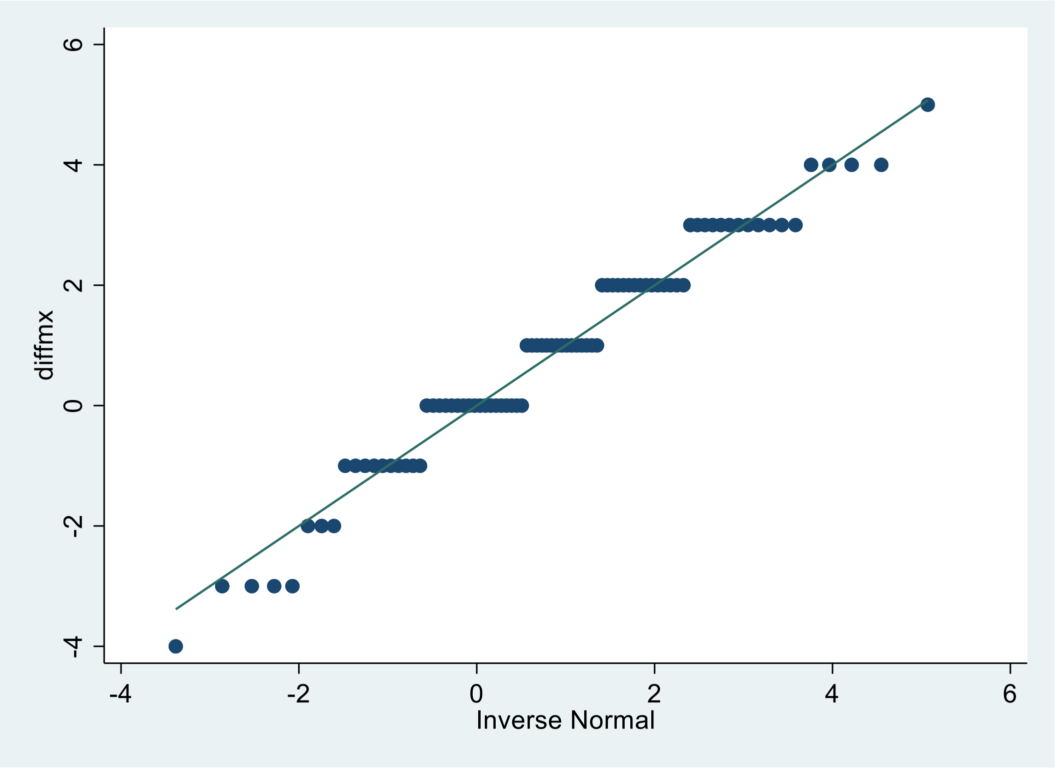


Figure 3. The quantiles of the normal distribution.

We also perform the Shapiro-Wilk test.

```
. swilk diffmx

                   Shapiro–Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
      diffmx |         83    0.99195      0.570    -1.235    0.89156
```

Based on the plots and the result of Shapiro-Wilk test, we can conclude that the variable is normally distributed. In addition, the number of observations = 83>30. Therefore, the the central limit theorem applies.

b) Since we are only interested in one variable, the difference in the degree of pain. The variable is normally distributed. A one-sample t test will be appropriate.

c) The hypothesis is given by: $H_0: \mu = 0$, $H_a: \mu \neq 0$.
The one-sample t test is performed in stata:

```
. ttest diffmx == 0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
  diffmx |      83    .8433735    .2053372    1.870711    .4348924    1.251855
------------------------------------------------------------------------------
    mean = mean(diffmx)                                           t =   4.1073
H0: mean = 0                                     Degrees of freedom =       82

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0001          Pr(T > t) = 0.0000
. display invttail(82,0.025)
1.9893186
```

Since t=4.1073 >1.9893, we will reject the null hypothesis and conclude that there is a difference in degree of pain during maximal activity while on Motrin compared to placebo.

d) The hypothesis is given by: $H_0: \mu = 0$, $H_a: \mu < 0$.
We calculate the two-sided 95% confidence interval.

```
. ci means diffmx

    Variable |        Obs        Mean    Std. err.       [95% conf. interval]
-------------+---------------------------------------------------------------
      diffmx |         83    .8433735    .2053372        .4348924    1.251855
```

Since the null value 0 is not within the two-sided 95% CI (0.43, 1.25), we reject the null hypothesis and conclude that Motrin is associated with a lower degree of pain during maximal activity compared to placebo.

Answer 8.139

We first calculate the average HgbA1c level across all visits.

```
. sort id

. by id: egen avg_a1c = mean(gly_a1c)
```

We generate a subset id for every visit and find the median of the average HgbA1c level.

```
. by id: gen sub_id=_n

. tabstat avg_a1c if sub_id ==1, statistics( median )

    Variable |       p50
-------------+----------
     avg_a1c |  8.659773
------------------------
```

We then split the boys into two groups: controlled (the average HgbA1c level < 8.66) and uncontrolled ((the average HgbA1c level >= 8.66).

```
. by id: gen cate =.
(910 missing values generated)

. replace cate=1 if avg_a1c < 8.66
(437 real changes made)

. replace cate=0 if avg_a1c >= 8.66
(473 real changes made)
```

We calculate the growth that is change in weight divided by change in age.

```
. by id: gen maxrow_id = _N

. by id: gen first_wt = wt_kg if sub_id == 1
(800 missing values generated)

. by id: gen first_wt2 = sum(first_wt)

. by id: gen last_wt = wt_kg if sub_id == maxrow_id
(800 missing values generated)
```

```
. gen wt_diff = last_wt - first_wt2
(800 missing values generated)

. by id: gen first_age = age_yrs if sub_id == 1
(800 missing values generated)

. by id: gen first_age2 = sum(first_age)

. by id: gen last_age = age_yrs if sub_id == maxrow_id
(800 missing values generated)

. gen age_diff = last_age - first_age2
(800 missing values generated)

. gen growth = wt_diff / age_diff
(805 missing values generated)
```

We assess the normality of the variable growth for both controlled and uncontrolled groups.

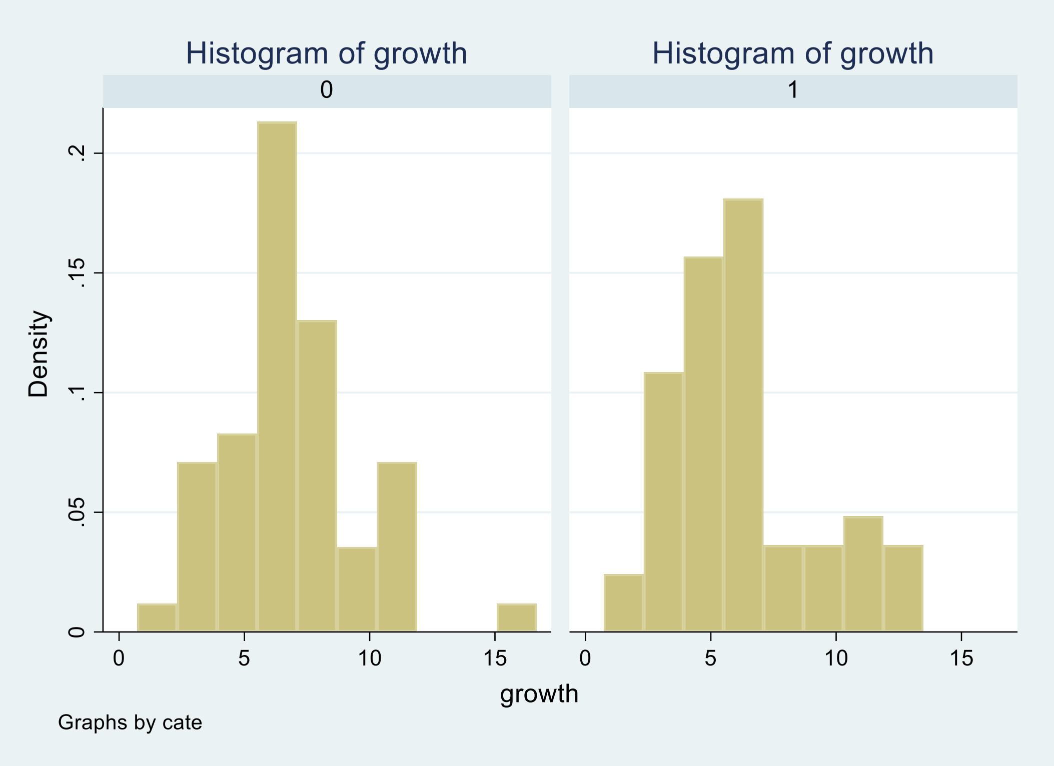


Figure 4. Histograms of growth for uncontrolled and controlled group.

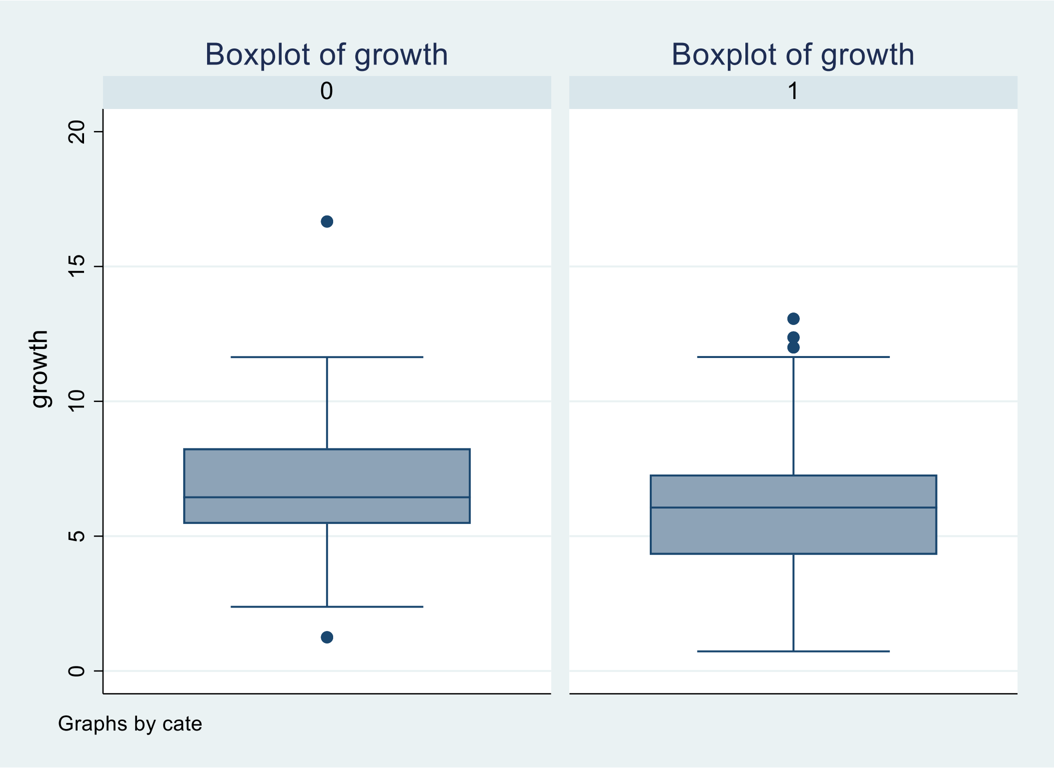
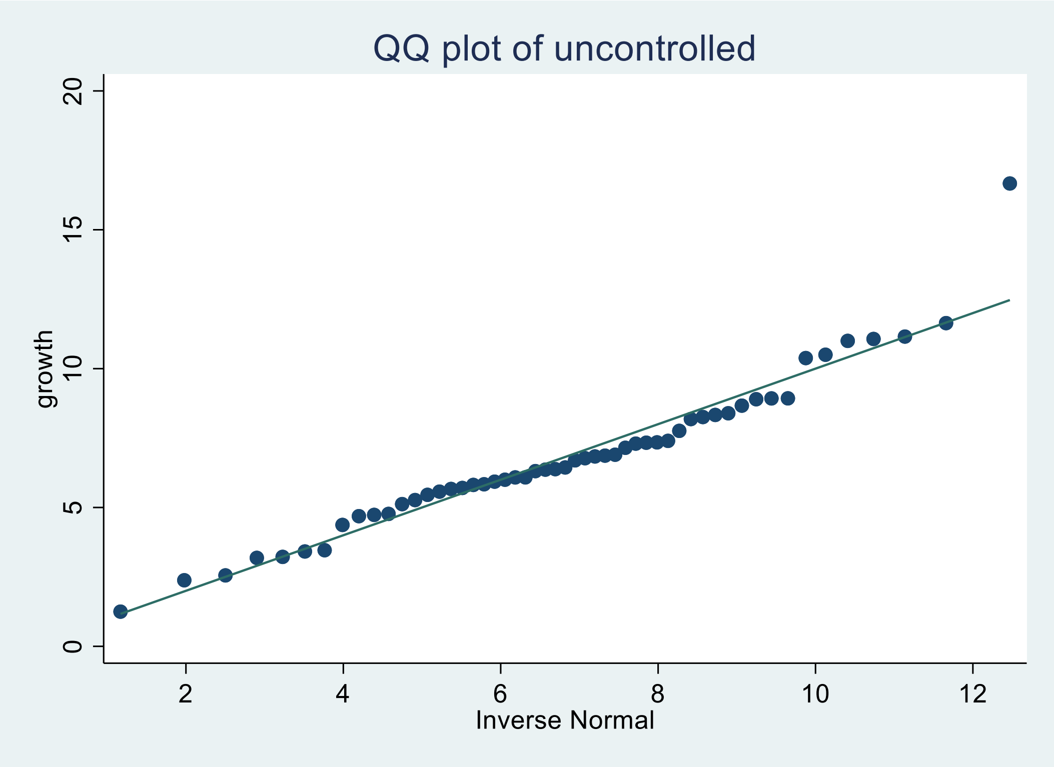


Figure 5. Boxplots of growth for uncontrolled and controlled group.



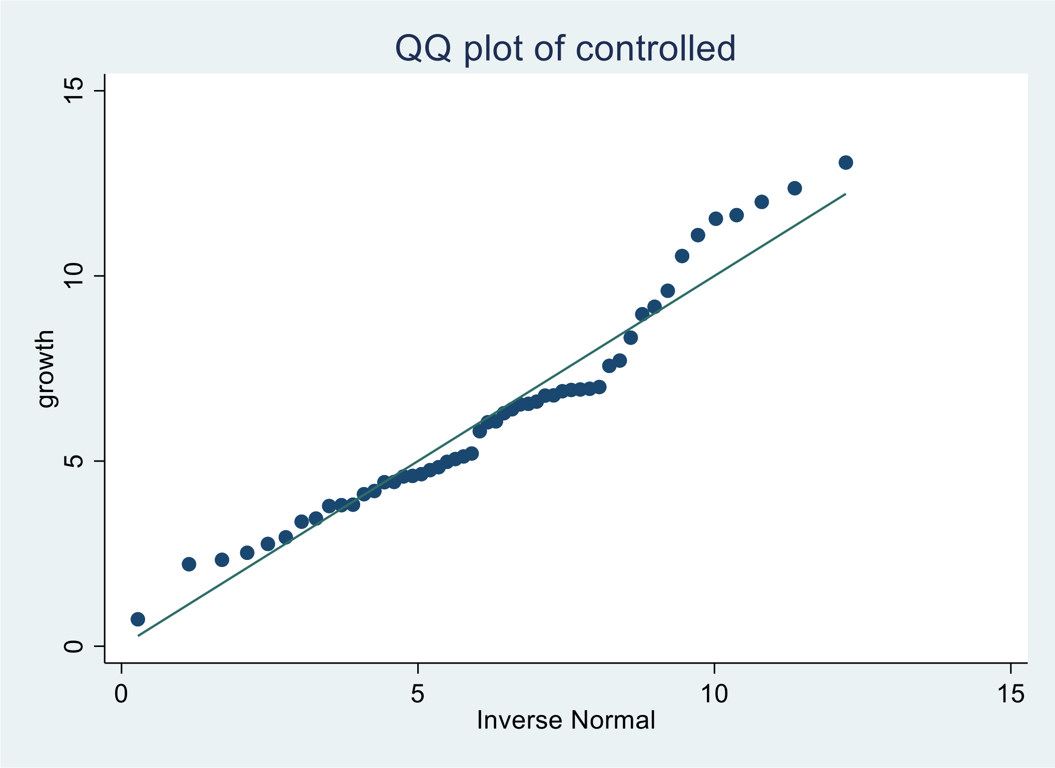


Figure 6. QQ plots of growth for uncontrolled and controlled group.

```
. swilk growth if cate == 0

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
      growth |         53    0.95108      2.409     1.881    0.02996

.  swilk growth if cate == 1

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
      growth |         52    0.94945      2.452     1.917    0.02760
```

The figures and results of Shapiro-Wilk tests show that the two groups largely conform to the normal distribution.

Then we assess the equality of variance between two groups.

```
. sdtest growth, by(cate)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      53    6.819387    .3722129    2.709751    6.072487    7.566287
       1 |      52    6.246733    .3985221    2.873783    5.446667    7.046799
---------+--------------------------------------------------------------------
Combined |     105    6.535787    .2726247    2.793572    5.995162    7.076412
------------------------------------------------------------------------------
    ratio = sd(0) / sd(1)                                         f =   0.8891
H0: ratio = 1                                    Degrees of freedom =   52, 51

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.3370         2*Pr(F < f) = 0.6741           Pr(F > f) = 0.6630
```

We conclude that the variance is equal. Therefore, we use a two-sample t test with equal variance to test the following hypothesis:

$$H_0: \mu_{uncontrolled} - \mu_{controlled} = 0, H_a: \mu_{uncontrolled} - \mu_{controlled} \neq 0$$

```
. ttest growth, by(cate)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      53    6.819387    .3722129    2.709751    6.072487    7.566287
       1 |      52    6.246733    .3985221    2.873783    5.446667    7.046799
---------+--------------------------------------------------------------------
Combined |     105    6.535787    .2726247    2.793572    5.995162    7.076412
---------+--------------------------------------------------------------------
    diff |            .5726539    .5450016               -.5082282    1.653536
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   1.0507
H0: diff = 0                                     Degrees of freedom =      103

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8521         Pr(|T| > |t|) = 0.2958          Pr(T > t) = 0.1479
```

Since the p-value is 0.2958 > 0.05, we cannot reject the null hypothesis. Therefore, we conclude that there is no difference in growth of the boys with controlled HgbA1c and uncontrolled HgbA1c.

Additional problem

a) We create a new variable to label the groups.

```
. gen expo=.
(124 missing values generated)

. replace expo=0 if lead_grp == 1
(78 real changes made)

. replace expo=1 if lead_grp == 2 | lead_grp == 3
(46 real changes made)
```

Let expo = 0 denote the unexposed group and expo = 1 denote the exposed group.

Then we plot the boxplots for both groups.

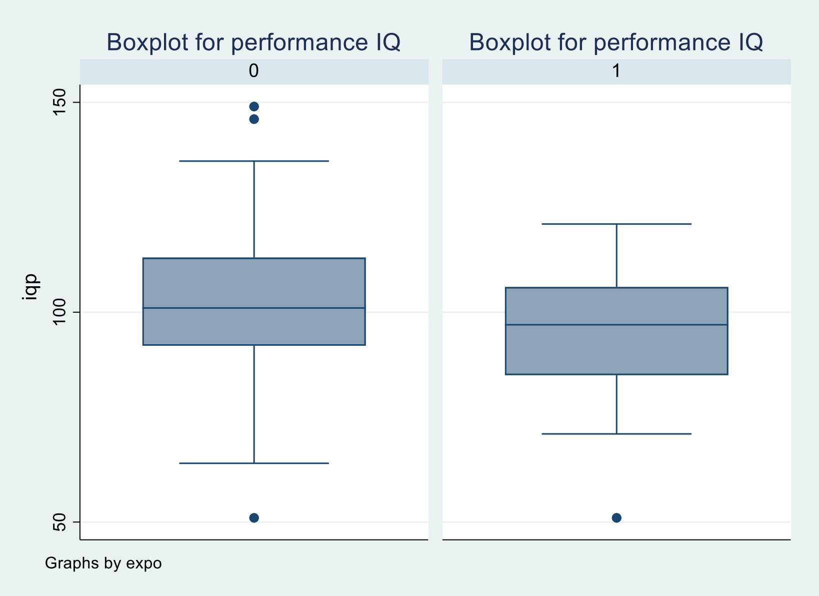


Figure 7. Boxplots of the unexposed group and the exposed group.

From the graph we can see that the unexposed group has three outliners and the exposed group has one outliner. We generate descriptive statistics for both groups.

```
. by expo, sort : summarize iqp

--------------------------------------------------------------------------------
-> expo = 0

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         iqp |         78    102.7051    16.78675         51        149

--------------------------------------------------------------------------------
-> expo = 1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         iqp |         46    94.93478    13.34733         51        121
```

From the boxplots and descriptive statistics, we can find that the two higher outliners of the unexposed group are approximately between 140 and 149. We can also discover that the lower outliners for both unexposed and exposed groups are 51. We display the ids of the values that satisfy those conditions.

```
. tab id expo if iqp > 140

           |     expo
        id |         0 |     Total
-----------+-----------+----------
       117 |         1 |         1
       139 |         1 |         1
-----------+-----------+----------
     Total |         2 |         2
```

```
. tab id expo if iqp == 51

           |          expo
        id |         0          1 |     Total
-----------+----------------------+----------
       135 |         1          0 |         1
       314 |         0          1 |         1
-----------+----------------------+----------
     Total |         1          1 |         2
```

b) Therefore, there are three outliners for the unexposed group. The ids of two higher outliers are 117 and 139. The id of the lower outliner is 135.
c) There is one outliner for the exposed group. The id of the outliner is 314.