Lara Haase
Lhaase
Assignment 1
Intermediate Statistics - 90-777

1.)

a.)

hw1

```
set type double
set more off
clear all

cd "C:\Users\Lara\Documents\Stats\Assignment 1"

import excel "C:\Users\Lara\Documents\Stats\Assignment 1\Water_Conservation_Data.xlsx", sheet("Sheet1") firstrow clear

summarize WATER_2006 SUMMER_07, detail

total TREAT1 TREAT2 TREAT3

count if GROUP==4

graph box SUMMER_07, over(GROUP)

sort GROUP
by GROUP: summarize SUMMER_07, detail

by GROUP: summarize WATER_2006, detail

graph twoway scatter SUMMER_07 WATER_2006, ytitle("Summer 2007 Usage ('000s)") xtitle("2006 Usage ('000s))") title("Water Usage from 2006 to 2007")

mean(SUMMER_07) if GROUP==4
mean(SUMMER_07) if GROUP==1
mean(SUMMER_07) if GROUP==2
mean(SUMMER_07) if GROUP==3


display binomial(135, 133, 0.9931)

```

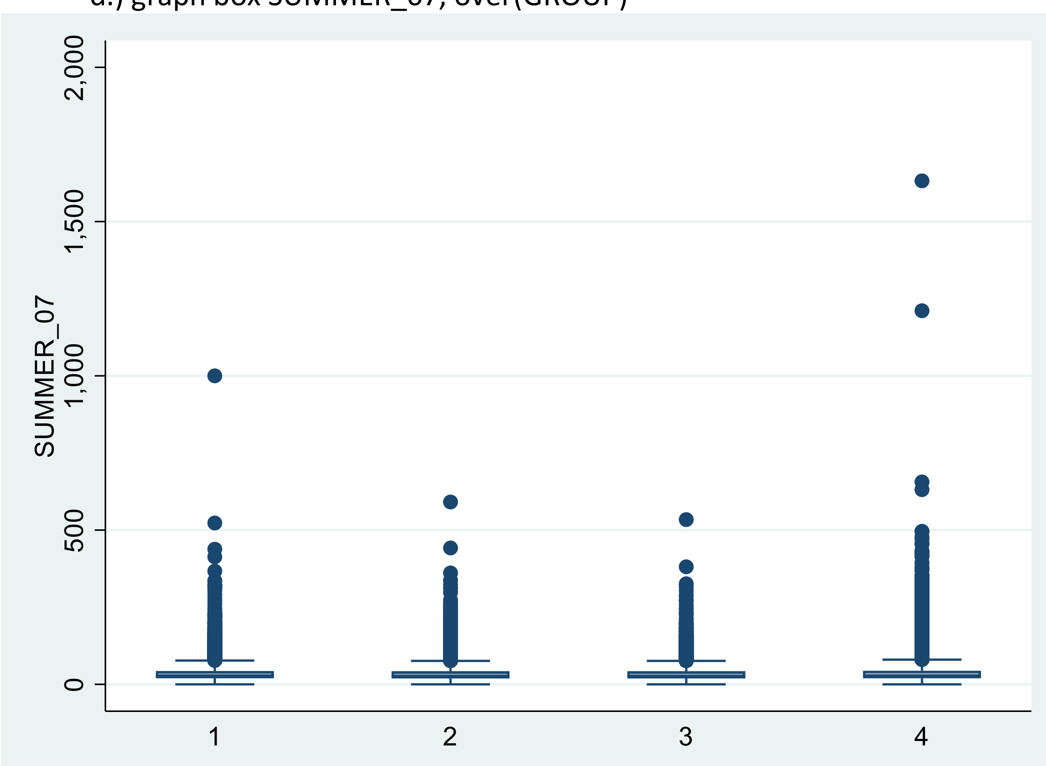
```
.
. summarize WATER_2006 SUMMER_07, detail

                         WATER_2006
-------------------------------------------------------------
      Percentiles      Smallest
 1%           23             20
 5%           25             20
10%           27             20       Obs             106,669
25%           33             20       Sum of Wgt.     106,669

50%           46                      Mean           58.31386
                        Largest       Std. Dev.      41.13629
75%           69           1000
90%          104           1000       Variance       1692.194
95%          133           1048       Skewness       5.402207
99%          211           2441       Kurtosis        139.449

                          SUMMER_07
-------------------------------------------------------------
      Percentiles      Smallest
 1%            3              0
 5%           11              0
10%           14              0       Obs             106,669
25%           20              0       Sum of Wgt.     106,669

50%           28                      Mean           36.16832
                        Largest       Std. Dev.      28.96305
75%           43            656
90%           68           1000       Variance       838.8581
95%           88           1211       Skewness       5.811943
99%          142           1632       Kurtosis       150.9183
```

b.)

c.)

```
.
. total TREAT1 TREAT2 TREAT3

Total estimation                  Number of obs   =    106,669

--------------------------------------------------------------
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
      TREAT1 |      11675    101.967      11475.15    11874.85
      TREAT2 |      11675    101.967      11475.15    11874.85
      TREAT3 |      11676   101.9708      11476.14    11875.86
--------------------------------------------------------------


.
. count if GROUP==4
  71,643
```

d.) graph box SUMMER_07, over(GROUP)



```
. by GROUP: summarize SUMMER_07, detail

-> GROUP = 1

                          SUMMER_07
-------------------------------------------------------------
      Percentiles      Smallest
 1%            3              0
 5%           11              0
10%           14              0       Obs              11,675
25%           20              0       Sum of Wgt.      11,675

50%           28                      Mean           36.35281
                        Largest       Std. Dev.       30.4252
75%           43            413
90%           69            438       Variance        925.693
95%           88            523       Skewness       5.991404
99%          147           1000       Kurtosis       112.3092
```

```
-> GROUP = 2

                          SUMMER_07
-------------------------------------------------------------
      Percentiles      Smallest
 1%            3              0
 5%           11              0
10%           14              0       Obs              11,675
25%           19              0       Sum of Wgt.      11,675

50%           27                      Mean           35.42741
                        Largest       Std. Dev.      28.13309
75%           42            337
90%           66            361       Variance       791.4707
95%           85            442       Skewness       3.702768
99%          145            591       Kurtosis       33.43722
```

```
-> GROUP = 3

                          SUMMER_07
-------------------------------------------------------------
      Percentiles      Smallest
 1%            4              0
 5%           11              0
10%           14              0       Obs              11,676
25%           19              0       Sum of Wgt.      11,676

50%           27                      Mean             34.858
                        Largest       Std. Dev.      26.33525
75%           42            322
90%           65            326       Variance       693.5454
95%           84            381       Skewness       3.506011
99%          132            534       Kurtosis       30.59214
```

```
-> GROUP = 4

                          SUMMER_07
-------------------------------------------------------------
      Percentiles      Smallest
 1%            3              0
 5%           11              0
10%           15              0       Obs              71,643
25%           20              0       Sum of Wgt.      71,643

50%           28                      Mean           36.47254
                        Largest       Std. Dev.      29.25198
75%           44            631
90%           68            656       Variance       855.6785
95%           89           1211       Skewness       6.342103
99%          142           1632       Kurtosis       186.4384
```

All of the treatment groups are positively(right) skewed because the means are greater than the medians.

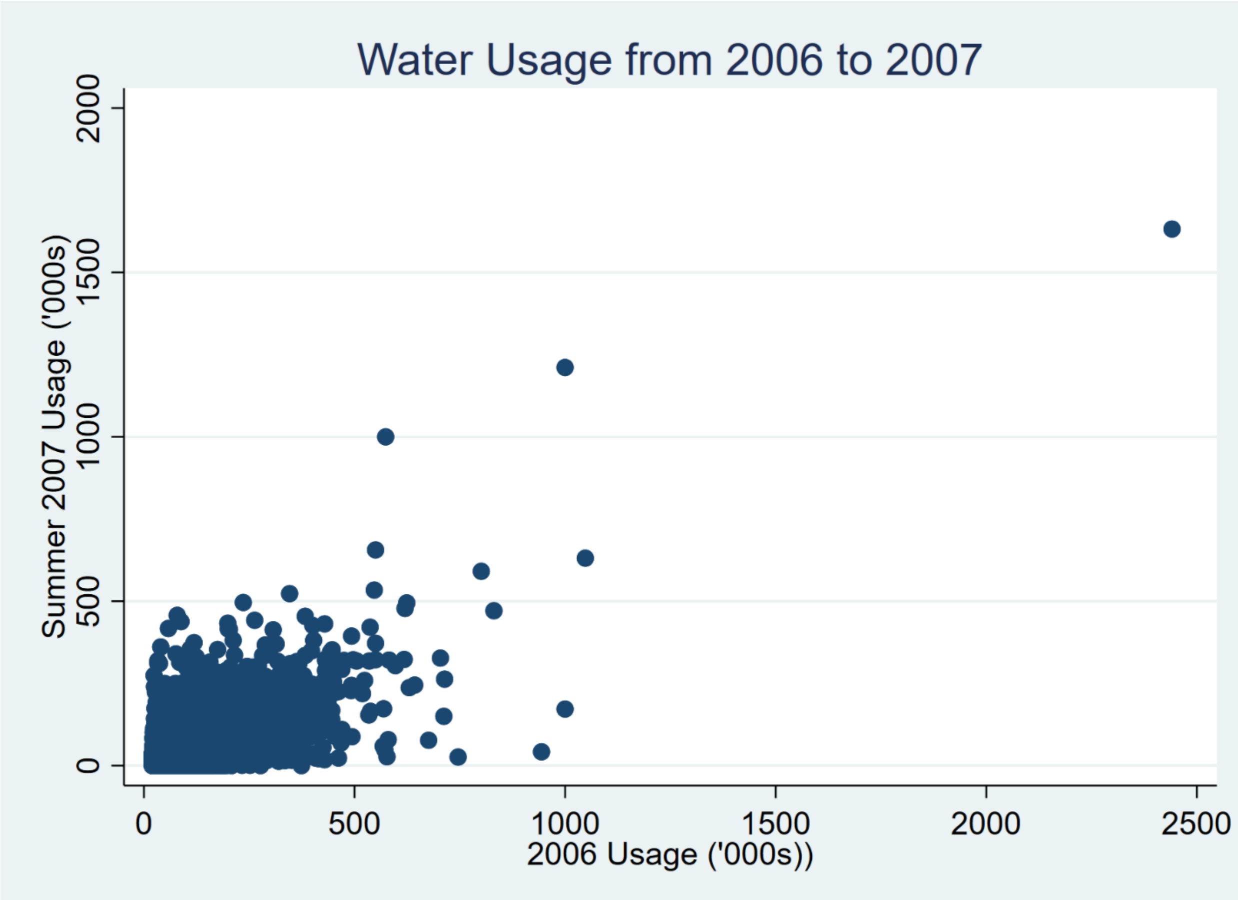
e.)

```
-> GROUP = 1

                          WATER_2006
-------------------------------------------------------------
      Percentiles      Smallest
 1%           23             20
 5%           25             20
10%           27             20       Obs              11,675
25%           33             20       Sum of Wgt.      11,675

50%           46                      Mean           58.42647
                        Largest       Std. Dev.      39.95928
75%           69            475
90%          104            574       Variance       1596.744
95%          131            580       Skewness        3.30767
99%          215            676       Kurtosis       24.37675

-> GROUP = 2

                          WATER_2006
-------------------------------------------------------------
      Percentiles      Smallest
 1%           23             20
 5%           25             20
10%           27             20       Obs              11,675
25%           33             20       Sum of Wgt.      11,675

50%           45                      Mean           58.17559
                        Largest       Std. Dev.      41.24874
75%           68            462
90%          103            550       Variance       1701.459
95%          134            714       Skewness       3.723298
99%          214            801       Kurtosis       31.86967
```

```
-> GROUP = 3

                           WATER_2006
-------------------------------------------------------------
      Percentiles      Smallest
 1%           23             20
 5%           25             20
10%           27             20       Obs              11,676
25%           34             20       Sum of Wgt.      11,676

50%           46                      Mean           58.43559
                        Largest       Std. Dev.      40.66821
75%           69            547
90%          104            597       Variance       1653.903
95%          132            746       Skewness       4.176852
99%          210           1000       Kurtosis       48.74561

-------------------------------------------------------------
-> GROUP = 4

                           WATER_2006
-------------------------------------------------------------
      Percentiles      Smallest
 1%           23             20
 5%           25             20
10%           27             20       Obs              71,643
25%           33             20       Sum of Wgt.      71,643

50%           46                      Mean            58.2982
                        Largest       Std. Dev.      41.38286
75%           69            944
90%          104           1000       Variance       1712.541
95%          133           1048       Skewness        6.16856
99%          211           2441       Kurtosis       186.7398
```

The mean of the Control group is 58.3, while the mean of treatment groups 1, 2 and 3 are 58.43, 58.18, and 58.44 respectively. The differences between groups are very close. The median of the Control group is 46, while the medians of groups 1, 2, and 3 are 46, 45 and 46. There are almost no differences here. This means that randomization was successful in balancing the mean and median water consumption between the treatment groups in 2006.

f.)



There appears to be a slightly positive relationship between the water usage in 2006 vs 2007. This may be supportive of the idea that the treatments have an affect on usage. I would expect usage from one year to be predictive of the next year's usage, which would create a strong positive correlation. However the lack of strength in the correlation may have been disturbed by the treatments. This can only be teased out if the correlation is examined between groups.

g.) If high water using household were assigned to the control group, it would dampen the possible positive outcomes on the treatments, because household that are already more conservative in their water usage would be the ones receiving the treatment, leaving less room for water usage reduction.

h.)

```
. mean(SUMMER_07) if GROUP==4

Mean estimation                    Number of obs   =     71,643

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   SUMMER_07 |   36.47254    .109287      36.25834    36.68674
--------------------------------------------------------------

. mean(SUMMER_07) if GROUP==1

Mean estimation                    Number of obs   =     11,675

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   SUMMER_07 |   36.35281   .2815821      35.80086    36.90475
--------------------------------------------------------------

. mean(SUMMER_07) if GROUP==2

Mean estimation                    Number of obs   =     11,675

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   SUMMER_07 |   35.42741   .2603688      34.91704    35.93778
--------------------------------------------------------------

. mean(SUMMER_07) if GROUP==3

Mean estimation                    Number of obs   =     11,676

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   SUMMER_07 |     34.858   .2437196      34.38027    35.33573
--------------------------------------------------------------
```

Treatment 1 difference= (control) 36.47 – 36.35 = 0.12 >> about 120 gallons less
Treatment 2 difference = (control) 36.47 – 35.43 = 1.04 >> about 1040 gallons less
Treatment 3 difference = (control) 36.47 – 34.86 = 1.61 >> about 1610 gallons less

Treatment 1 % difference = (0.12/36.47)*100 = 0.329%
Treatment 2 % difference = (1.04/36.47)*100 = 2.852 %
Treatment 3 % difference = (1.61/36.47)*100 = 4.415%

Treatment 3 appears to have the strongest effect, though Treatment 2 also appears to have some effect. The statistical significance of these differences would have to be calculated to determine if these differences are not simply due to chance.

2.)

a.) p^x*q^(n-x)

= (144/145)^135

= 0.393

or 39.3%

b.) Yes, I believe the independence assumption is reasonable in this circumstance. Each launch of the shuttle requires a set of rigorous checks and tests to assure that the equipment is prepared for the extreme conditions of the launches. Theses test should allow for the technicians to correct for any errors or flaws, so that the equipment is at the same standards of condition at the beginning of each launch, making the chances of each launch success independent of the previous launches.

c.)

```
end of do-file

. do "C:\Users\Lara\AppData\Local\Temp\STDd68_000000.tmp"

. display binomial(135, 133, 0.9931)
.23897302


.
end of do-file


.
```