Topics in Biostatistics **Homework 2** **Due Tuesday 21st January at 2.30pm**

Directions: Answer the following two exercises, showing all relevant work. Conclusions and justifications are to be given in clear detailed English. For each exercise, clearly write down all needed assumptions/requirements. Type up your solutions or write ***very*** neatly.

1. Extracorporeal membrane oxygenation (ECMO) is a potentially life-saving procedure that is used to treat newborn babies who suffer from severe respiratory failure. An experiment was conducted in which 20 babies were treated with ECMO and 30 babies were treated with conventional medical therapy (CMT). At the end of the study, 11 of the CMT babies died (19 survived), and only 2 of the ECMO babies died (18 survived).

   (a) Test whether these data suggest that the therapies significantly differ. Also, test the one-tailed alternative that the ECMO procedure is better than (in terms of better survival rates) the conventional therapy.

   (b) Find and interpret the Odds Ratio (OR) of survival comparing the ECMO therapy with the CMT, and provide a 95% confidence interval for the true OR. Hint: you must do this part by hand; see the Chapter 1 Notes.

   (c) Let's now alter the above data by supposing that of the 20 ECMO babies, only one died (19 survived). Explain why the usual (chi-square) test statistic is inappropriate here, and analyze these "new" data using the correct statistical analysis. For this part, perform only the one-tailed test specified as the second test in part (a) of this exercise.

2. [Data taken from Walker, 2002:61] A drug company is developing a new appetite suppressing compound for use in weight reduction. A preliminary study of 35 obese patients provided the following data on patients' body weights (in pounds) before ("PreW", in pounds) and after ("PostW", in pounds) 10 weeks of treatment with the new compound. These data are graphed on the next page (red squares) with two additional points (black circles): one at the point (100,100) and one at the point (300,300); these latter two points are not a part of the data, but may be useful for part (c) of this exercise.

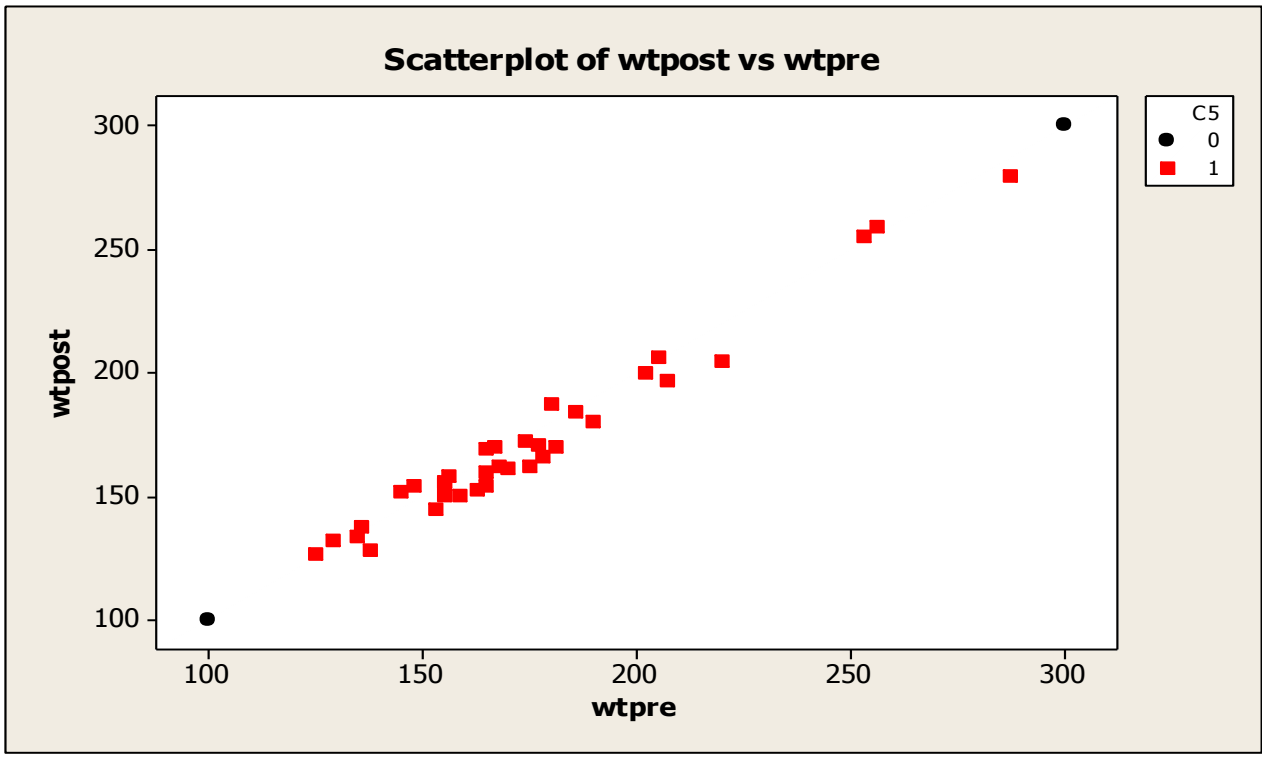
Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
PreW	165	202	256	155	135	175	180	174	136	168	207	155	220	163	159	253	138	287
PostW	160	200	259	156	134	162	187	172	138	162	197	155	205	153	150	255	128	280
Sub	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	
PreW	177	181	148	167	190	165	155	153	205	186	178	129	125	165	156	170	145	
PostW	171	170	154	170	180	154	150	145	206	184	166	132	127	169	158	161	152	

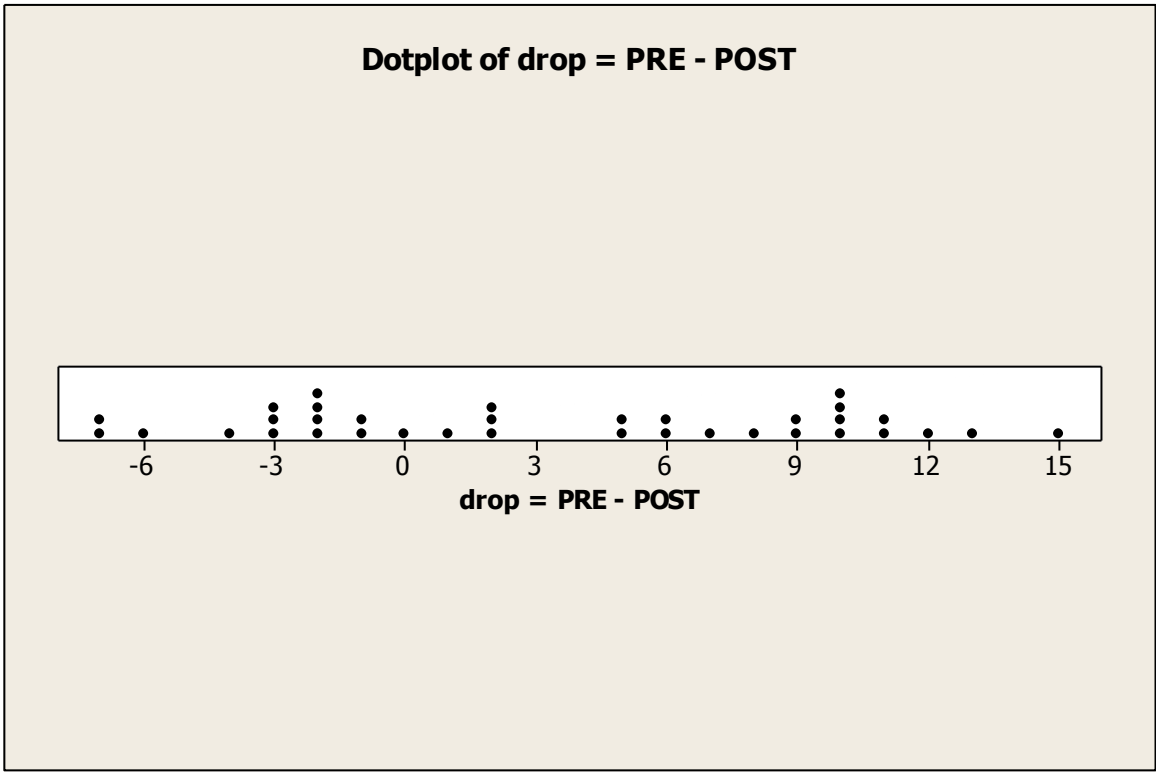
   (a) Does the new treatment look at all promising? Be specific and list all necessary assumptions and/or reasons why some usual one(s) are not needed here.

   (b) Does a subjects' "Pre" weight appear to be a good linear predictor of his/her "Post" weight? Again, be specific and **list all necessary assumptions and/or reasons why some usual assumption(s) are not needed here**.

   (c) Reconcile the analyses in parts (a) and (b). That is, discuss any connection(s) (if any) between the two analyses.

Graphs for Exercise 2





Output A

Chi-Square Test: CMT, ECMO

```
Expected counts are printed below observed counts

          ECMO      CMT    Total
    1       18       19       37
          14.80    22.20

    2        2       11       13
           5.20     7.80
Total       20       30       50

Chi-Sq =  0.692 +  0.461 +
          1.969 +  1.313 = 4.435
DF = 1, P-Value = 0.035
```

Output B

Paired T-Test and CI: wtpre, wtpost

```
Paired T for wtpre - wtpost

                 N      Mean     StDev   SE Mean
wtpre           35    174.94     35.94      6.07
wtpost          35    171.49     35.45      5.99
Difference      35      3.46      6.34      1.07

95% lower bound for mean difference: 1.65
T-Test of mean difference = 0 (vs > 0): T-Value = 3.23  P-Value = 0.001
```

Output C

Regression Analysis: wtpost versus wtpre

```
The regression equation is
wtpost = 1.61 + 0.971 wtpre

Predictor        Coef     SE Coef          T        P
Constant        1.615       5.407       0.30    0.767
wtpre         0.97101     0.03030      32.05    0.000

S = 6.348       R-Sq = 96.9%     R-Sq(adj) = 96.8%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1       41397       41397   1027.31    0.000
Residual Error    33        1330          40
Total             34       42727

Unusual Observations
Obs      wtpre     wtpost         Fit      SE Fit    Residual    St Resid
  3        256     259.00      250.19        2.68        8.81        1.53 X
 18        287     280.00      280.29        3.56       -0.29       -0.06 X
X denotes an observation whose X value gives it large influence.
```