PHP 2500 Introduction to Biostatistics

Problem Set Six Solutions

1/
Chapter 11, #5

a.
The two samples are paired since we are taking repeated observations on the same subject.

b.
H_0: $\mu_{CF} - \mu_{OB} = 0$ versus H_0: $\mu_{CF} - \mu_{OB} \neq 0$

c.
```
ttest cf=ob

Paired t test

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      cf |      14    4.443571    .2589319    .9688344    3.884183     5.00296
      ob |      14    4.080714    .2824898     1.05698    3.470432    4.690997
---------+--------------------------------------------------------------------
    diff |      14    .3628571    .1084984    .4059638    .1284606    .5972537
------------------------------------------------------------------------------

                     Ho: mean(cf - ob) = mean(diff) = 0

  Ha: mean(diff) < 0         Ha: mean(diff) ~= 0        Ha: mean(diff) > 0
       t =   3.3444               t =   3.3444              t =   3.3444
   P < t =   0.9974          P > |t| =   0.0053           P > t =   0.0026
```

The p-value for the two sided test is .0053.

d.
Therefore, we reject the null hypothesis that the difference in means is 0.

2/
Chapter 11, #7

a.
```
ttest after12=after24, level(90)

Paired t test

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [90% Conf. Interval]
---------+--------------------------------------------------------------------
 after12 |       7    69.85714    15.95593    42.21543    38.85189    100.8624
 after24 |       7    30.42857    7.981697    21.11758     14.9187    45.93845
---------+--------------------------------------------------------------------
    diff |       7    39.42857    11.86603    31.39457    16.37073    62.48641
------------------------------------------------------------------------------

                Ho: mean(after12 - after24) = mean(diff) = 0

  Ha: mean(diff) < 0         Ha: mean(diff) ~= 0        Ha: mean(diff) > 0
       t =   3.3228               t =   3.3228              t =   3.3228
   P < t =   0.9920          P > |t| =   0.0159           P > t =   0.0080
```

Our one sided 95% CI for $\delta=\mu_{12}-\mu_{24}$ is $\delta\geq 16.36$.

b.
Our test of the null hypothesis yields a one-sided p-value of .0080 therefore, we reject the null hypothesis at the 5% level of significance.

3/
Chapter 11, #9

a.
Assuming equal variances,

```
ttesti 23 111 8 24 109 8,  level(99)

Two-sample t test with equal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [99% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      23         111    1.668115           8     106.298     115.702
       y |      24         109    1.632993           8    104.4156    113.5844
---------+--------------------------------------------------------------------
combined |      47    109.9787    1.163542     7.97684    106.8523    113.1052
---------+--------------------------------------------------------------------
    diff |                   2    2.334368               -4.278482    8.278482
------------------------------------------------------------------------------
Degrees of freedom: 45

                      Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =   0.8568              t =   0.8568              t =   0.8568
   P < t =   0.8019          P > |t| =   0.3961          P > t =   0.1981
```

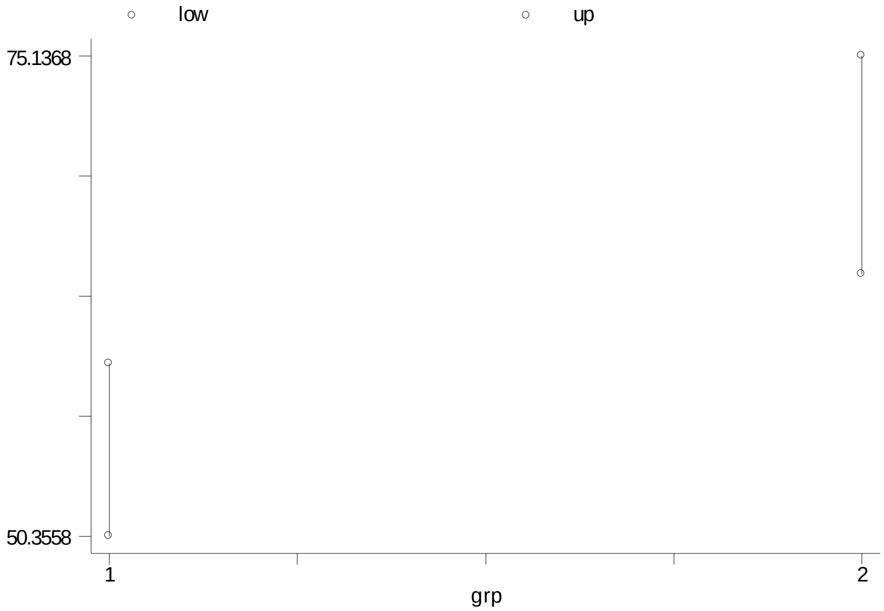
The two-sided hypothesis test yields a p-value of .3961 therefore we cannot reject the null hypothesis that the two populations of women have the same mean arterial blood pressure.

b. 99% CI is (-4.28, 8.28). This interval contains the value 0 which was to be expected since we did not reject the null hypothesis that the difference in mean arterial bp between the two groups of women was 0.

4/
Chapter 11, #10

a.

```
ttesti 156 54.8 28.1 148 69.5 34.7

Two-sample t test with equal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |     156        54.8    2.249801        28.1    50.35577    59.24423
       y |     148        69.5    2.852322        34.7    63.86315    75.13685
---------+--------------------------------------------------------------------
combined |     304    61.95658     1.85161    32.28392    58.31294    65.60022
---------+--------------------------------------------------------------------
    diff |               -14.7    3.612929               -21.8097   -7.590298
------------------------------------------------------------------------------
Degrees of freedom: 302
```



Since the Confidence interval on the difference does not contain zero, there appears to be evidence that the populations means are different. Note that it is tempting to simply notice that the two confidence intervals do not overlap and the two population means are not likely to be equal, but this is not the correct way to assess if there is a difference or not.

b.

```
ttesti 156 54.8 28.1 148 69.5 34.7

Two-sample t test with equal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |     156        54.8    2.249801        28.1    50.35577    59.24423
       y |     148        69.5    2.852322        34.7    63.86315    75.13685
---------+--------------------------------------------------------------------
combined |     304    61.95658     1.85161    32.28392    58.31294    65.60022
---------+--------------------------------------------------------------------
    diff |               -14.7    3.612929                -21.8097   -7.590298
------------------------------------------------------------------------------
Degrees of freedom: 302

                      Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -4.0687               t =  -4.0687            t =  -4.0687
   P < t =   0.0000          P > |t| =   0.0001          P > t =   1.0000
```

The p-value for the two sided test is .0001, therefore I reject the null hypothesis that the population means are equal.

c.
A 95% CI for the true difference in population means is (-21.8097, -7.59028)

d.

```
. ttesti 156 172.5 68.8 148 185.5 69.0

Two-sample t test with equal variances

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

```
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |     156       172.5    5.508408        68.8    161.6188    183.3812
       y |     148       185.5    5.671765          69    174.2913    196.7087
---------+--------------------------------------------------------------------
combined |     304    178.8289    3.962633    69.09087    171.0312    186.6267
---------+--------------------------------------------------------------------
    diff |                 -13    7.905814               -28.55746    2.557458
------------------------------------------------------------------------------
Degrees of freedom: 302

                      Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -1.6444               t =  -1.6444              t =  -1.6444
   P < t =   0.0506         P > |t| =   0.1011          P > t =   0.9494
```

The p-value for this two sided test is .1011, therefore I cannot reject that the two groups of husbands have the same mean carbohydrate intake.

5/
Chapter 11, #13

a.
```
summ  bed80  bed86

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
       bed80 |      51    4.686275   1.009756          3          7
       bed86 |      51    4.235294   1.176235          2          8
```

b.
```
ttest  bed80= bed86, unpaired

Two-sample t test with equal variances

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   bed80 |      51    4.686275    .1413942    1.009756    4.402276    4.970273
   bed86 |      51    4.235294    .1647059    1.176235    3.904473    4.566116
---------+--------------------------------------------------------------------
combined |     102    4.460784    .1103036    1.114011    4.241972    4.679597
---------+--------------------------------------------------------------------
    diff |            .4509804    .2170722                .0203153    .8816454
------------------------------------------------------------------------------
Degrees of freedom: 100

                  Ho: mean(bed80) - mean(bed86) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =   2.0776               t =   2.0776              t =   2.0776
   P < t =   0.9798         P > |t| =   0.0403          P > t =   0.0202
```

The p-value for the two sided test is .0403, therefore I would reject the null hypothesis that the mean number of beds in the two years was the same.

c.
```
ttest  bed80= bed86

Paired t test

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   bed80 |      51    4.686275    .1413942    1.009756    4.402276    4.970273
   bed86 |      51    4.235294    .1647059    1.176235    3.904473    4.566116
---------+--------------------------------------------------------------------
    diff |      51    .4509804      .08075    .5766706    .2887892    .6131716
------------------------------------------------------------------------------
```

```
                  Ho: mean(bed80 - bed86) = mean(diff) = 0

  Ha: mean(diff) < 0         Ha: mean(diff) ~= 0        Ha: mean(diff) > 0
       t =   5.5849                t =   5.5849              t =   5.5849
   P < t =   1.0000          P > |t| =   0.0000          P > t =   0.0000
```

The p-value for the paired test is ~0, therefore I would again reject the null hypothesis that the mean number of beds was the same in each year.

d. The difference between the two tests is that the paired t-test recognizes (and takes in to account) that we have repeated observations on the same "unit". So in a sense we have controlled a great deal of variability. Treating the data as if it came from two independent populations is dangerous in that we ignore the natural pairing in the data. In this case we happen to reach the same conclusion.

e.

```
gen diff=bed80-bed86

ci diff

    Variable |     Obs         Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
        diff |      51     .4509804      .08075        .2887892    .6131716
```

95% CI for the true difference is (.289, .613)

6/
Chapter 11, #15

a.

```
ttest  pdi, by ( trtment) unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      73    91.91781    1.929775      16.488    88.07087    95.76474
       1 |      70    97.77143    1.755225    14.68527    94.26985     101.273
---------+--------------------------------------------------------------------
combined |     143    94.78322    1.325531    15.85104    92.16289    97.40354
---------+--------------------------------------------------------------------
    diff |            -5.85362     2.60861                -11.0109   -.6963371
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  140.247

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -2.2440                t =  -2.2440              t =  -2.2440
   P < t =   0.0132          P > |t| =   0.0264          P > t =   0.9868
```

The p-value for the two sided test is .0264. Therefore, we reject the null hypothesis that the mean pdi score is the same for children's in both treatment groups.

b.

```
. ttest  mdi, by ( trtment) unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      74    103.1622    1.914019    16.46501    99.34753    106.9768
       1 |      70       106.4    1.741754    14.57256    102.9253    109.8747
---------+--------------------------------------------------------------------
```

```
combined |     144    104.7361    1.300364    15.60437    102.1657    107.3065
---------+--------------------------------------------------------------------
    diff |           -3.237838     2.58789               -8.353798    1.878122
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  141.386

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =  -1.2511               t =  -1.2511              t =  -1.2511
   P < t =   0.1065          P > |t| =   0.2129          P > t =   0.8935
```

The p-value for this two sided test is .2129. Therefore, we cannot reject the null hypothesis that the mean mdi score is the same for children in both treatment groups.

c.
These tests suggest that the type of surgical treatment will impact a child's psychomotor development but not their mental development.

7/
Chapter 13, #7

a.
```
signtest  time1= time2

Sign test

        sign |    observed    expected
-------------+------------------------
    positive |           9           7
    negative |           5           7
        zero |           0           0
-------------+------------------------
         all |          14          14

One-sided tests:
  Ho: median of time1 - time2 = 0 vs.
  Ha: median of time1 - time2 > 0
      Pr(#positive >= 9) =
         Binomial(n = 14, x >= 9, p = 0.5) =  0.2120

  Ho: median of time1 - time2 = 0 vs.
  Ha: median of time1 - time2 < 0
      Pr(#negative >= 5) =
         Binomial(n = 14, x >= 5, p = 0.5) =  0.9102

Two-sided test:
  Ho: median of time1 - time2 = 0 vs.
  Ha: median of time1 - time2 ~= 0
      Pr(#positive >= 9 or #negative >= 9) =
         min(1, 2*Binomial(n = 14, x >= 9, p = 0.5)) =  0.4240
```

The p-value for the two sided test is .4240 therefore, we cannot reject the null hypothesis that the median difference in respiratory rates for the two times is equal to 0.

b.
```
signrank  time1= time2

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        9        80.5        52.5
    negative |        5        24.5        52.5
        zero |        0           0           0
-------------+---------------------------------
         all |       14         105         105
```

```
unadjusted variance      253.75
adjustment for ties       -0.12
adjustment for zeros       0.00
                      ----------
adjusted variance        253.63

Ho: time1 = time2
             z =   1.758
    Prob > |z| =   0.0787
```

If we use the Wilcoxon Sign-Rank test we find a two-sided p-value of .0787. Therefore, I would not reject the null hypothesis that the median respiratory rate at the two times is equal.

c.
Using the two-sided test, we do reach the same conclusions.

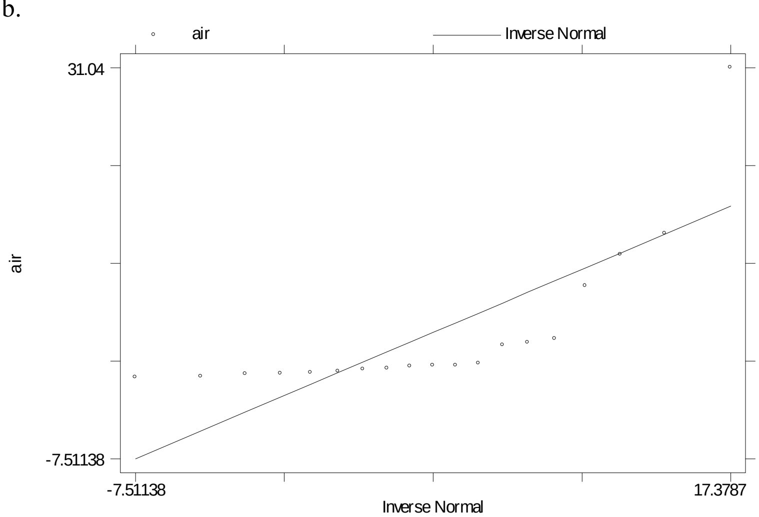
8/
Chapter 13, #8

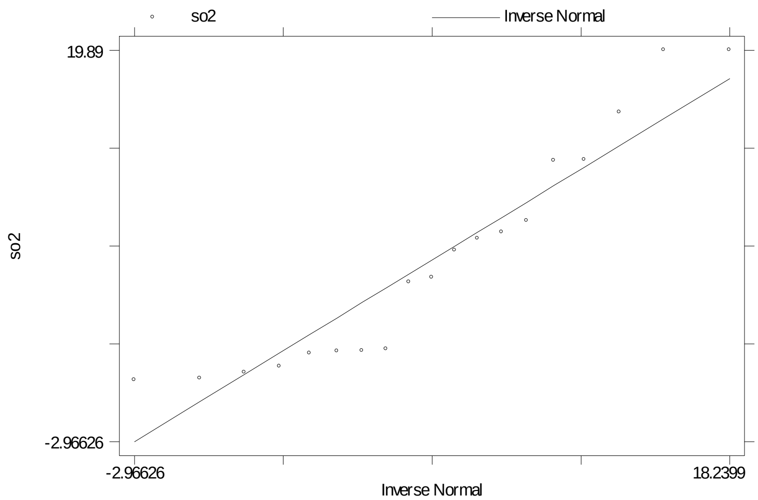
a.

```
signrank  air= so2

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        5          43          95
    negative |       14         147          95
        zero |        0           0           0
-------------+---------------------------------
         all |       19         190         190

unadjusted variance      617.50
adjustment for ties        0.00
adjustment for zeros       0.00
                      ----------
adjusted variance        617.50

Ho: air = so2
             z =  -2.093
    Prob > |z| =   0.0364
```

The p-value for the two-sided test is .0364. Therefore, I would reject the null hypothesis that the median difference is equal to zero.

b.





Neither of the variables are normally distributed, therefore a ttest would not have been appropriate.

9/
Chapter 13, #11

a.

```
ranksum  age, by(sex)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

         sex |      obs    rank sum    expected
-------------+---------------------------------
           0 |       16         221         224
           1 |       11         157         154
-------------+---------------------------------
    combined |       27         378         378

unadjusted variance      410.67
adjustment for ties       -0.38
                     ----------
adjusted variance        410.29

Ho: age(sex==0) = age(sex==1)
             z =  -0.148
    Prob > |z| =   0.8823
```

The p-value for the rank sum test is .8823, therefore, I cannot reject the null hypothesis that the median age is the same for males and females.

b.

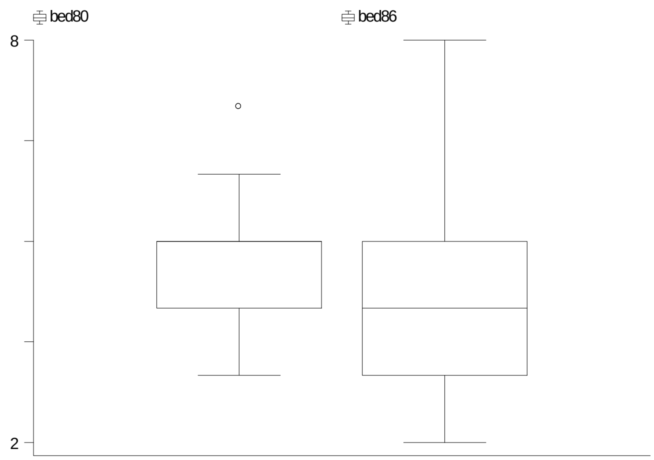
```
. by sex: centile age, centile(50)

_______________________________________________________________________________
-> sex = 0

                                                       -- Binom. Interp. --
    Variable |     Obs  Percentile      Centile        [95% Conf. Interval]
-------------+-------------------------------------------------------------
         age |      16          50         82.5        68.31055    123.6895

_______________________________________________________________________________
-> sex = 1

                                                       -- Binom. Interp. --
    Variable |     Obs  Percentile      Centile        [95% Conf. Interval]
-------------+-------------------------------------------------------------
         age |      11          50           87        58.85091    141.4691
```

```
. by sex: summ age

______________________________________________________________________________
-> sex = 0

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------
         age |      16       97.25     39.4656         46        175

______________________________________________________________________________
-> sex = 1

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------
         age |      11    107.6364     66.1321         53        277
```

Notice that for each sex, the median age ≠ mean age which implies the data is not symmetric. Therefore, a ttest would not be appropriate.

10/
Chapter 13, #13

a.



b.

```
signrank  bed80= bed86

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       25         950         513
    negative |        2          76         513
        zero |       24         300         300
-------------+---------------------------------
         all |       51        1326        1326

unadjusted variance    11381.50
adjustment for ties     -409.50
adjustment for zeros   -1225.00
                     ----------
adjusted variance       9747.00

Ho: bed80 = bed86
             z =   4.426
    Prob > |z| =   0.0000
```

The p-value is ~0 and thus I would reject the null hypothesis that the median number of beds is the same in both years.

c.

```
. ranksum  bed, by( year)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        year |      obs    rank sum    expected
-------------+---------------------------------
          80 |       51        2963      2626.5
          86 |       51        2290      2626.5
-------------+---------------------------------
    combined |      102        5253        5253

unadjusted variance    22325.25
adjustment for ties    -1817.57
                     ----------
adjusted variance      20507.68

Ho: bed(year==80) = bed(year==86)
             z =   2.350
    Prob > |z| =   0.0188
```

The p-value is .0188. Therefore, I would reject the null hypothesis that the median number of beds is the same in both years.

d.
The sign rank test assumes the data is paired whereas the rank sum test assumes we have two independent populations. In this example, both tests yield the same conclusions. The sign-rank test would be the appropriate choice here since the data is actually paired (repeated measurements on the same state).

e.
Non-parametric tests yielded the same results. This is actually not surprising since from part a we see the data is approximately symmetric (which would imply the mean=median).

11/
Chapter 14, #9

a.

```
. cii 45 4, level(90)

                                                         -- Binomial Exact --
    Variable |     Obs        Mean    Std. Err.       [90% Conf. Interval]
-------------+-------------------------------------------------------------
             |      45     .0888889    .0424232         .0309353    .1919756
```

Therefore a 90% CI for the true proportion is (.031, .192)

b.
H_0: p=.22
H_a: p$\neq$ .22

c.

```
bitesti 45 4 .22

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
       45          4          9.9       0.22000      0.08889

  Pr(k >= 4)            = 0.994273  (one-sided test)
  Pr(k <= 4)            = 0.018872  (one-sided test)
  Pr(k <= 4 or k >= 17) = 0.030768  (two-sided test)
```

d.

The p-value is .03, therefore we would reject the null hypothesis that the true proportion is the same for children in the special ed program as for all 3rd graders.

e.

```
sampsi .22 .1, onesample power(.95) alpha(.05)

Estimated sample size for one-sample comparison of proportion
  to hypothesized value

Test Ho: p = 0.2200, where p is the proportion in the population

Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9500
 alternative p =   0.1000

Estimated required sample size:

             n =      119
```

Therefore, a sample of size 119 would be required.

12/
Chapter 14, #11

a.
For prepaid plan, p^=13/311 = .042
For Medicaid plan, p^ = 22/310 = .071

b.

```
. prtesti 311 13 310 22, level(90) count

Two-sample test of proportion                      x: Number of obs =      311
                                                   y: Number of obs =      310

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       z     P>|z|       [90% Conf. Interval]
---------+--------------------------------------------------------------------
       x |  .0418006    .0113485   3.68336   0.0002        .023134    .0604673
       y |  .0709677    .0145836   4.86627   0.0000       .0469798    .0949556
---------+--------------------------------------------------------------------
    diff | -.0291671    .0184789                          -.0595622     .001228
         |  under Ho:   .0185087  -1.57586   0.1151
------------------------------------------------------------------------------

               Ho: proportion(x) - proportion(y) = diff = 0

       Ha: diff < 0            Ha: diff ~= 0            Ha: diff > 0
        z = -1.576               z = -1.576               z = -1.576
    P < z =  0.0575        P > |z| =  0.1151          P > z =  0.9425
```

c.
The p-value for the two sided test is .1151, therefore, I cannot reject the null hypothesis that the proportions for the two populations are identical (at the 10% significance level).

13/
Chapter 14, #12

a.
p= 97(.247)+161(.174) / (97+161)
 = .20

b.

```
prtesti 97 .247 161 .174

Two-sample test of proportion                     x: Number of obs =       97
                                                  y: Number of obs =      161

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       .247    .0437885    5.64075   0.0000       .1611761    .3328239
       y |       .174     .029878    5.82368   0.0000       .1154402    .2325598
---------+--------------------------------------------------------------------
    diff |       .073    .0530106                           -.030899     .176899
         |  under Ho:    .0515516    1.41606   0.1568
------------------------------------------------------------------------------

               Ho: proportion(x) - proportion(y) = diff = 0

       Ha: diff < 0            Ha: diff ~= 0             Ha: diff > 0
        z =  1.416               z =  1.416                z =  1.416
    P < z =  0.9216        P > |z| =  0.1568           P > z =  0.0784
```

c.
The p-value for the two sided test is .1568. Therefore, I cannot reject the null hypothesis that the proportion of intravenous drug users who have a positive tuberculin skin test result are the same for those who share needles and those who do not.

d.
95% CI for the true difference in proportions is (-0.03, .18).

14/
Chapter 15, #8

a.
```
tabi 1250 991\1387 1666, chi2

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |      1250        991 |      2241
         2 |      1387       1666 |      3053
-----------+----------------------+----------
     Total |      2637       2657 |      5294

          Pearson chi2(1) =  55.3553   Pr = 0.000
```

Therfore, $\chi^2 = 55.35$

b.
The chi-square value we found is significantly different than 1, thus we would conclude that the behavior of college students changes from one year to the next.

c.
```
prtesti 2637 1250 2657 991, count

Two-sample test of proportion                     x: Number of obs =     2637
                                                  y: Number of obs =     2657

------------------------------------------------------------------------------
Variable |       Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |   .4740235    .0097236    48.7497   0.0000       .4549656    .4930815
       y |    .372977    .0093818    39.7553   0.0000        .354589    .3913651
---------+--------------------------------------------------------------------
    diff |   .1010465    .0135117                           .0745639     .127529
         |  under Ho:    .0135813    7.44011   0.0000
------------------------------------------------------------------------------
```

```
              Ho: proportion(x) - proportion(y) = diff = 0

       Ha: diff < 0             Ha: diff ~= 0             Ha: diff > 0
        z =  7.440                z =  7.440                z =  7.440
    P < z =  1.0000         P > |z| =  0.0000          P > z =  0.0000
```

Again – we would reject the null hypothesis that the proportion of students who drove while drinking is the same in the two years.

d.
95% CI for true difference in proportions is (`.0745639, .127529`).

15/
Chapter 15, #11

a.

```
tabi 2040 367 327\149 60 48\288 25 70\703 197 252\425 62 88\121 72 79, chi2

           |               col
       row |         1          2          3 |     Total
-----------+---------------------------------+----------
         1 |      2040        367        327 |      2734
         2 |       149         60         48 |       257
         3 |       288         25         70 |       383
         4 |       703        197        252 |      1152
         5 |       425         62         88 |       575
         6 |       121         72         79 |       272
-----------+---------------------------------+----------
     Total |      3726        783        864 |      5373

         Pearson chi2(10) = 209.0933   Pr = 0.000
```

Since our chi-square statistic is significantly different than 10, we would conclude that the results are not consistent across studies.

b.
One problem might be that we might report a direction of effect that is the opposite of the true association.

16/
Chapter 15, #13

a.

```
mcci 27 20 12 68

                 | Controls               |
Cases            |   Exposed   Unexposed  |     Total
-----------------+------------------------+----------
         Exposed |        27          20  |        47
       Unexposed |        12          68  |        80
-----------------+------------------------+----------
           Total |        39          88  |       127

McNemar's chi2(1) =      2.00    Prob > chi2 = 0.1573
Exact McNemar significance probability       = 0.2153

Proportion with factor
        Cases       .3700787
        Controls    .3070866     [95% Conf. Interval]
                   ---------     --------------------
        difference  .0629921     -.0314928    .157477
        ratio       1.205128      .9301778   1.561351
        rel. diff.  .0909091     -.0292189    .211037

        odds ratio  1.666667      .7759952   3.739198   (exact)
```

b.

Since McNemar's chi-square statistic is not significantly different than 1, we can not reject the null hypothesis of no association between retirement status and cardiac arrest.

c.

```
mcci 27 20 12 68

                 | Controls               |
Cases            |   Exposed   Unexposed  |     Total
-----------------+------------------------+----------
         Exposed |        27          20  |        47
       Unexposed |        12          68  |        80
-----------------+------------------------+----------
           Total |        39          88  |       127

McNemar's chi2(1) =      2.00    Prob > chi2 = 0.1573
Exact McNemar significance probability       = 0.2153

Proportion with factor
        Cases       .3700787
        Controls    .3070866     [95% Conf. Interval]
                   ---------     --------------------
        difference  .0629921     -.0314928    .157477
        ratio       1.205128      .9301778   1.561351
        rel. diff.  .0909091     -.0292189    .211037

        odds ratio  1.666667      .7759952   3.739198   (exact)
```

OR= 1.67

d.

95% CI for population OR is (.776, 3.739). Yes this interval contains the value 1 which was to be expected since we did not reject the null hypothesis of no association.

17/

Chapter 15, #15

a.

```
mcci  2 8 2 33

                 | Controls               |
Cases            |   Exposed   Unexposed  |     Total
-----------------+------------------------+----------
         Exposed |         2           8  |        10
       Unexposed |         2          33  |        35
-----------------+------------------------+----------
           Total |         4          41  |        45

McNemar's chi2(1) =      3.60    Prob > chi2 = 0.0578
Exact McNemar significance probability       = 0.1094

Proportion with factor
        Cases       .2222222
        Controls    .0888889     [95% Conf. Interval]
                   ---------     --------------------
        difference  .1333333      -.020997   .2876637
        ratio            2.5      .9382946   6.661021
        rel. diff.  .1463415      .0066704   .2860125

        odds ratio         4      .7982264    38.6707   (exact)
```

b. We see that McNemar's chi-square statistic takes on the value 3.60 which is not significantly different than 1. Therefore, we cannot reject the null hypothesis of no association (but we must do so with caution. An inference based entirely on p-values could be questionable here).

18/

Chapter 15, #16

```
cci 108 163 117 268
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+----------------------
           Cases |       108         163  |        271      0.3985
        Controls |       117         268  |        385      0.3039
-----------------+------------------------+----------------------
           Total |       225         431  |        656      0.3430
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
      Odds ratio |         1.517697       |  1.080447    2.130222  (exact)
 Attr. frac. ex. |          .341107       |  .0744574    .5305654  (exact)
 Attr. frac. pop |         .1359393       |
                 +-----------------------------------------------
                              chi2(1) =     6.32  Pr>chi2 = 0.0119
```

a.
Relative Odds for Smokers vs Non-Smokers = 1.517697

b.
95% CI for population OR:
(1.08, 2.13)

c.
Chi-Square Statistic is 6.32 which is significantly different than 1 therefore we reject the null hypothesis of no association.

19/
Chapter 15, #17

a.

```
. cci 28 251 6 273, level(99)
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+----------------------
           Cases |        28         251  |        279      0.1004
        Controls |         6         273  |        279      0.0215
-----------------+------------------------+----------------------
           Total |        34         524  |        558      0.0609
                 |                        |
                 |      Point estimate    |  [99% Conf. Interval]
                 |------------------------+----------------------
      Odds ratio |         5.075697       |  1.592784    22.07797  (exact)
 Attr. frac. ex. |         .8029827       |  .3721686     .954706  (exact)
 Attr. frac. pop |         .0805861       |
                 +-----------------------------------------------
                              chi2(1) =    15.16  Pr>chi2 = 0.0001
```

b.
OR = 5.076

c.
99% CI for population OR: (1.59, 22.08)

20/
Chapter 15, #18

a.

P(abortion|0 alcohol) = 6793/33164 = .205
P(abortion|1-2 drinks) = 2068/9099 = .227
P(abortion|3-6 drinks)=776/3069 = .253

P(abortion|7-20 drinks) = 456/1527 = .299
P(abortion|21+ drinks)=98/287 = .341

b.

```
 tabodds case alcohol [fweight=freq],or base(1)

---------------------------------------------------------------------------
     alcohol |  Odds Ratio       chi2       P>chi2     [95% Conf. Interval]
-------------+-------------------------------------------------------------
           1 |    1.000000          .            .            .           .
           2 |    1.141822      21.71       0.0000      1.079823   1.207381
           3 |    1.313780      39.20       0.0000      1.205875   1.431340
           4 |    1.652876      77.69       0.0000      1.476384   1.850467
           5 |    2.012933      32.47       0.0000      1.574664   2.573183
---------------------------------------------------------------------------
Test of homogeneity (equal odds): chi2(4)  =   118.91
                                  Pr>chi2  =   0.0000

Score test for trend of odds:     chi2(1)  =   152.17
                                  Pr>chi2  =   0.0000
```

- Note that the relative odds for each alcohol "level" are in the column labeled Odds Ratio.

c.
The 95% CI's are highlighted in the above table.

d.
None of the 95% OR's contain the value 1 suggesting that spontaneous abortions and alcohol use are associated. Also, all of the estimated OR's >1 and increase with increasing alcohol consumption.