PHP 2500 Introduction to Biostatistics

Problem Set Five Solutions

1/
Chapter 9, #5

a.
```
cii 10 130 11.8
```

```
    Variable |     Obs         Mean    Std. Err.       [95% Conf. Interval]
-------------+------------------------------------------------------------
             |      10          130    3.731488        121.5588    138.4412
```

95% CI for the population mean is (121.6, 138.4).

b.
We are 95% confident that the interval (121.6, 138.4) covers the true mean systolic blood pressure.

c.
```
. cii 10 84 9.1, level(90)
```

```
    Variable |     Obs         Mean    Std. Err.       [90% Conf. Interval]
-------------+------------------------------------------------------------
             |      10           84    2.877673         78.7249     89.2751
```

90% CI for the population mean is (78.7 , 89.3).

d.
```
cii 10 84 9.1, level(99)
```

```
    Variable |     Obs         Mean    Std. Err.       [99% Conf. Interval]
-------------+------------------------------------------------------------
             |      10           84    2.877673        74.64804    93.35196
```

99% CI for the population mean is (74.6 , 93.4).

e.
The 99% CI is wider than the 90% CI for the true mean diastolic blood pressure.

2/
Chapter 9, #8

a.
```
cii 12 4.49 .83
```

```
    Variable |     Obs         Mean    Std. Err.       [95% Conf. Interval]
-------------+------------------------------------------------------------
             |      12         4.49    .2396004        3.962643    5.017357
```

95% CI for the true population mean FVC is (3.962643,5.017357).

b.
```
cii 12 4.49 .83, level(90)
```

```
    Variable |     Obs         Mean    Std. Err.       [90% Conf. Interval]
-------------+------------------------------------------------------------
             |      12         4.49    .2396004        4.059705    4.920295
```

90% CI for the true population mean FVC is (4.059705,4.920295).

c.

```
cii 12 3.71 .62
```

```
    Variable |     Obs         Mean    Std. Err.       [95% Conf. Interval]
-------------+--------------------------------------------------------------
             |      12         3.71    .1789786        3.316071    4.103929
```

95% CI for the true population mean FEV1 is (3.316071,4.103929).

d.
We are assuming that the underlying distributions are Normal.

3/
Chapter 9, #9

a.

```
cii 14 29.6 3.6
```

```
    Variable |     Obs         Mean    Std. Err.       [95% Conf. Interval]
-------------+--------------------------------------------------------------
             |      14         29.6    .9621405        27.52142    31.67858
```

95% CI for the true population mean is (27.52, 31.68).

b.
length= 31.68 – 27.52 = 4.16

c.
Find n such that $2(1.96)(\sigma/\sqrt{n})=3$
→ $\sqrt{n}=2(1.96)(3.6)/3$
→ n= 22.13
Therefore, take n=23

d.
Find n such that $2(1.96)(\sigma/\sqrt{n})=2$
→ $\sqrt{n}=2(1.96)(3.6)/2$
→ n= 49.79
Therefore, take n=50

4/
Chapter 9, #10

a.

```
ci  ibw
```

```
    Variable |     Obs         Mean    Std. Err.       [95% Conf. Interval]
-------------+--------------------------------------------------------------
         ibw |      18     112.7778     3.39988        105.6047    119.9509
```

95% CI for true mean ideal body weight is (105.6, 120.0).

b.
No – the confidence interval does not contain the value 100. This implies that the population of insulin-dependent diabetics weighs on average more than their ideal body weight.

5/
Chapter 9 , #11

a.

```
ci  calcium, level(90)
```

```
    Variable |     Obs        Mean    Std. Err.       [90% Conf. Interval]
-------------+--------------------------------------------------------------
     calcium |       8       3.1425    .1805522        2.80043     3.48457
```

Therefore, the lower bound of the 95% one-sided interval is 2.80.

b.

```
. ci   albumin, level(90)

    Variable |     Obs        Mean    Std. Err.       [90% Conf. Interval]
-------------+--------------------------------------------------------------
     albumin |       8       40.375       1.068       38.35159    42.39841
```

Therefore, the lower bound of the 95% one-sided interval is 38.4.

c.
I believe that these patients have normal blood albumin levels (as the lower bound is in the normal range). However, the lower bound for calcium level is below normal levels.

6/
Chapter 14, #8

a.
phat = 473/488 = .97

b.

```
cii 488 473

                                                    -- Binomial Exact --
    Variable |     Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+--------------------------------------------------------------
             |     488     .9692623    .0078135        .9498101    .9826969
```

95% CI for the true population proportion is (.9498,.9827)

c.
We are 95% confident that the interval (.9498,.9827) covers the true population proportion p.

d.

```
cii 488 473,level(90)

                                                    -- Binomial Exact --
    Variable |     Obs        Mean    Std. Err.       [90% Conf. Interval]
-------------+--------------------------------------------------------------
             |     488     .9692623    .0078135        .9530638    .9809597
```

90% CI for the true population proportion is (.953, .981)

e.
The 90% CI is more narrow than the 95% CI.

7/
Chapter 14, # 10

a.

```
prtesti 114 11 96 7, count

Two-sample test of proportion                     x: Number of obs =      114
                                                  y: Number of obs =       96

------------------------------------------------------------------------------
Variable |      Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
```

```
---------+--------------------------------------------------------------------
       x |  .0964912     .027654   3.48923   0.0005       .0422904     .150692
       y |  .0729167    .0265361   2.74783   0.0060       .0209069    .1249265
---------+--------------------------------------------------------------------
    diff |  .0235746    .0383263                        -.0515437    .0986928
         |  under Ho:   .0387783   .607931   0.5432
```

a.
Estimate of p1-p2 = .0235746

b.
95% CI for the true difference in proportions is (-.05, .10).

d.
Since the 95% CI contains the value 0, I am 95% confident that the doctor's advice was not effective.

8/
Chapter 14,#11

a.

```
prtesti 311 13 310 22,count

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

------------------------------------------------------------------------------
Variable |      Mean    Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |  .0418006    .0113485   3.68336   0.0002        .019558    .0640433
       y |  .0709677    .0145836   4.86627   0.0000       .0423844    .0995511
---------+--------------------------------------------------------------------
    diff | -.0291671    .0184789                        -.0653851    .0070509
         |  under Ho:   .0185087  -1.57586   0.1151
------------------------------------------------------------------------------
```

Estimated Proportion of Patients who had visited a crisis center for those on prepaid plan = .0418

Estimated Proportion of Patients who had visited a crisis center for those on medicaid = .0710

b.

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

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

prtesti 311 13 310 22,count level(99)

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

------------------------------------------------------------------------------
Variable |      Mean    Std. Err.       z     P>|z|       [99% Conf. Interval]
---------+--------------------------------------------------------------------
       x |  .0418006    .0113485   3.68336   0.0002       .0125688    .0710325
       y |  .0709677    .0145836   4.86627   0.0000       .0334029    .1085326
---------+--------------------------------------------------------------------
    diff | -.0291671    .0184789                        -.0767656    .0184315
```

```
         |  under Ho:   .0185087  -1.57586   0.1151
```

- 90% CI for the true difference in proportions is (-.0596, .0012)
- 95% CI for the true difference in proportions (-0.065, .007)
- 99% CI for the true difference in proportions is (-0.0768, .0184)

c.
Since all of these intervals contain the value 0, I would conclude with 99% confidence that there is no difference in the proportion of people who visit a community crisis center for those on prepaid plan versus those on Medicaid.

9/
Chapter 11, #11

a.
H_0: $\mu_{nonsmpkers} = \mu_{smpkers}$
Ha: $\mu_{nonsmpkers} < \mu_{smpkers}$

b.

```
. ttesti 121 1.3 1.3 75 4.1 2.0, unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |     121         1.3    .1181818         1.3    1.066008    1.533992
       y |      75         4.1    .2309401           2    3.639842    4.560158
---------+--------------------------------------------------------------------
combined |     196    2.371429    .1501613    2.102258     2.07528    2.667577
---------+--------------------------------------------------------------------
    diff |                -2.8     .259423               -3.313961   -2.286039
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  113.052

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

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t = -10.7932                 t = -10.7932                 t = -10.7932
   P < t =   0.0000           P > |t| =   0.0000           P > t =   1.0000
```

The p-value for the one-sided test is ~0, therefore, we reject the null hypothesis that the difference in means is 0.

A 95% CI for the true difference in means between non-smokers and smokers is (`-3.313961,-2.286039`). Notice the interval does not contain 0. This too supports our conclusion that the means are different.

10/
Chapter 11,#12

```
ttesti 37 27.9 5.6 19 38.8 21.7, unequal

Two-sample t test with unequal variances

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      37        27.9    .9206343         5.6    26.03287    29.76713
       y |      19        38.8    4.978321        21.7    28.34093    49.25907
---------+--------------------------------------------------------------------
combined |      56    31.59821    1.898091    14.20401    27.79435    35.40207
```

```
---------+--------------------------------------------------------------------
    diff |               -10.9    5.062732              -21.48744    -.312555
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:   19.241

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

     Ha: diff < 0               Ha: diff ~= 0               Ha: diff > 0
       t =  -2.1530               t =  -2.1530                t =  -2.1530
   P < t =   0.0221           P > |t| =   0.0442           P > t =   0.9779
```

A 95% CI for the true difference in means is (-21.5, -.31). This CI does not contain the value 0 and so we reject the null hypothesis that the population means are equal (at the 5% level). Notice that two-sided p-value from the above ttest would yield the same conclusion.

11/
Chapter 11,#10

a.

```
ttesti 156 54.8 28.1 148 69.5 34.7, unequal

Two-sample t test with unequal 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    3.632815              -21.85077   -7.549226
------------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  282.944

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

     Ha: diff < 0               Ha: diff ~= 0               Ha: diff > 0
       t =  -4.0464               t =  -4.0464                t =  -4.0464
   P < t =   0.0000           P > |t| =   0.0001           P > t =   1.0000
```

A 95% CI for the true mean daily fat intake for husbands in the intervention group is (50.36, 59.24).
A 95% CI for the true mean daily fat intake for husbands in the control group is (63.86, 75.14).

Since the intervals do not overlap, it is unlikely that the population means of the two groups are equal.
c.
The 95% CI for the difference in mean daily fat intake for the two groups of husbands is (-21.85, -7.54). Since the interval does not contain 0, we reject the null hypothesis that the mean daily fat intake is the same for both groups of husbands (at the 5% level).

d.

```
Two-sample t test with unequal 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     7.90642              -28.55885    2.558845
```

A 95% CI for the true difference in mean carbohydrate intake for husbands in the two treatment groups is (-28.6, 2.6). Since this interval contains the value 0, we cannot reject (at the 5% level) the null hypothesis that the means are the same.

12/

length	z-score	confidence	p	sqrt(p(1-p))	n	**sample size**
0.1	1.645	90	0.55	0.4974937	267.8965	**268**
0.1	1.645	90	0.2	0.4	173.1856	**174**
0.1	1.645	90	0.9	0.3	97.4169	**97**
0.1	1.96	95	0.55	0.4974937	380.3184	**381**
0.1	1.96	95	0.2	0.4	245.8624	**246**
0.1	1.96	95	0.9	0.3	138.2976	**139**
0.1	2.58	99	0.55	0.4974937	658.9836	**659**
0.1	2.58	99	0.2	0.4	426.0096	**427**
0.1	2.58	99	0.9	0.3	239.6304	**240**
0.05	1.645	90	0.55	0.4974937	1071.586	**1072**
0.05	1.645	90	0.2	0.4	692.7424	**693**
0.05	1.645	90	0.9	0.3	389.6676	**390**
0.05	1.96	95	0.55	0.4974937	1521.274	**1522**
0.05	1.96	95	0.2	0.4	983.4496	**984**
0.05	1.96	95	0.9	0.3	553.1904	**554**
0.05	2.58	99	0.55	0.4974937	2635.934	**2636**
0.05	2.58	99	0.2	0.4	1704.038	**1705**
0.05	2.58	99	0.9	0.3	958.5216	**959**
0.025	1.645	90	0.55	0.4974937	4286.344	**4287**
0.025	1.645	90	0.2	0.4	2770.97	**2771**
0.025	1.645	90	0.9	0.3	1558.67	**1559**
0.025	1.96	95	0.55	0.4974937	6085.094	**6086**
0.025	1.96	95	0.2	0.4	3933.798	**3934**
0.025	1.96	95	0.9	0.3	2212.762	**2213**
0.025	2.58	99	0.55	0.4974937	10543.74	**10544**
0.025	2.58	99	0.2	0.4	6816.154	**6817**
0.025	2.58	99	0.9	0.3	3834.086	**3835**

13/ See attached stata log for this problem.

a. There is not a gender difference in either systolic blood pressure or heart rate at admission. In both cases, the p-values were quite large (p=0.4870 and 0=0.5784), and both 95% CIs on the gender difference included zero: -12.8 to 6.13 for sys and -9.9 to 5.54 for hra.

b. Yes. Patients with chronic renal failure die at a rate of 42.1% versus 17.7% for those without. The p-value testing equal probability of death was p=0.0113 and the confidence interval on the difference in the probability of death was 47% to 1.5%.

c. Yes. Patients who have surgery in the ICU are less likely to die (13.1% versus 27.9%) and the p-value testing equality was 0.0086 and the confidence interval for the difference was 3.8% to 25.9%.

d. T-tests that assume equal variances give very close answers to the test for the difference in proportions. This is because the test for the difference in proportions (in stata!) uses the null assumption that the proportions are equal and pools to estimate that common proportion. Notice that when w t-test is preformed assuming unequal variances, the answer is quite different (for CRN it goes from significant to not-significant at the 5% level). This is a very important point. Make sure you understand what is going on here!