Unit 7 – Stata for Analysis of One & Two Samples
Homework

SOLUTIONS

```
. *
. *****    1.)  Immediate command for one sample t test
. *****    ttesti n xbar sigma nullmean

. ttesti 12 1.2 0.6 1.0

One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      12         1.2    .1732051          .6    .8187782    1.581222
------------------------------------------------------------------------------
    mean = mean(x)                                                t =   1.1547
Ho: mean = 1.0                                 degrees of freedom =       11

   Ha: mean < 1.0               Ha: mean != 1.0                 Ha: mean > 1.0
 Pr(T < t) = 0.8637         Pr(|T| > |t|) = 0.2727          Pr(T > t) = 0.1363
```

INTERPRETATION:
In this sample of n=12, the observed mean was 1.2. A two-sided t-test of the null hypothesis that μ = 1.0 using sample standard deviation s = 0.6, yielded a t-statistic value of 1.15 and associated two sided p-value = 0.27. This is not statistically significant. The null hypothesis is not rejected. Conclude that these data do not provide statistically significant evidence that the mean serum creatinine among patients taking the new antibiotic μ ≠ 1.0.

```
. *
. ******    2.)  Immediate command for one sample, continuous outcome, 95% CI for mean
. ******    cii n xbar s, level(#)

. cii means 12 1.2 0.6, level(95)

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         12         1.2    .1732051        .8187782    1.581222
```

INTERPRETATION:
Based on this sample of n=12, with 95% confidence, the unknown mean serum creatinine among patients taking the new antibiotic is estimated to be between 0.82 and 1.58.

```
. *
. ******  3.)  Immediate command for one sample, binary outcome, test of event prob
. ******    EXACT test has command:  bitesti n #successes nullpi
. bitesti 5000 15 .005

       N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
    5000         15           25       0.00500      0.00300

  Pr(k >= 15)            = 0.987761  (one-sided test)
  Pr(k <= 15)            = 0.022046  (one-sided test)
  Pr(k <= 15 or k >= 36) = 0.044220  (two-sided test)
```

INTERPRETATION:
Assumption of the null hypothesis has led to an unlikely outcome (p-value = .04). Reject the null hypothesis that the event probability π = .005

```
. **     NORMAL APPROXIMATION test command:  prtesti n #successes nullpi, count

. prtesti 5000 15 .005, count

One-sample test of proportion                      x: Number of obs =     5000
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       .003   .0007734                      .0014841    .0045159
------------------------------------------------------------------------------
    p = proportion(x)                                             z =  -2.0050
Ho: p = 0.005

    Ha: p < 0.005               Ha: p != 0.005                 Ha: p > 0.005
 Pr(Z < z) = 0.0225         Pr(|Z| > |z|) = 0.0450          Pr(Z > z) = 0.9775
```

INTERPRETATION:
Among the 5000 men in the year 2000 cohort who were followed for one year, there were 15 new cases of MI, or 0.3%. A two-sided test of the null hypothesis that $\pi = .005$, using both exact and normal approximation approaches, yielded p-values of .04 and .045, respectively. This is marginally statistically significant. Conclude that these data provide marginally statistically significant evidence that the one year incidence of MI in the 2000 cohort $\pi \neq 0.005$.

```
. *
. ******     4.)  Immediate command for one sample, binary outcome, test of event prob
. ******     EXACT test has command:  bitesti n #successes nullpi

. bitesti 300 110 .30

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
      300         110           90     0.30000      0.36667

  Pr(k >= 110)            = 0.007765  (one-sided test)
  Pr(k <= 110)            = 0.994479  (one-sided test)
  Pr(k <= 70 or k >= 110) = 0.013875  (two-sided test)

. **    NORMAL APPROXIMATION test has command:  prtesti n #successes nullpi, count

. prtesti 300 110 .30, count

One-sample test of proportion                      x: Number of obs =      300
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .3666667   .0278222                      .3121362    .4211972
------------------------------------------------------------------------------
    p = proportion(x)                                             z =   2.5198
Ho: p = 0.3

     Ha: p < 0.3                Ha: p != 0.3                  Ha: p > 0.3
 Pr(Z < z) = 0.9941         Pr(|Z| > |z|) = 0.0117          Pr(Z > z) = 0.0059
```

INTERPRETATION:
Among the 300 women taking erythromycin regularly during weeks 24-28 of their pregnancies, there were 110 events of nausea complaints, or 36.7%. A two-sided test of the null hypothesis that $\pi = .30$, using both exact and normal approximation approaches, yielded a p-value of.01 in both instances. This is statistically significant. Reject the null hypothesis. Conclude that these data provide statistically significant evidence that the probability of nausea complaints among women taking erythromycin during weeks 24-28 of their pregnancy $\pi \neq 0.30$.

```
. *
. *****  5-8.) Following assumes you have downloaded data to desktop. Use sepsis.dta.

. use "/Users/cbigelow/Desktop/sepsis.dta", clear

. codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
id	455	455	228	1	455	Patient ID
treat	455	2	.4923077	0	1	Treatment
race	455	3	.4263736	0	2	Race
apache	454	38	15.3304	0	41	Baseline APACHE Score
o2del	168	168	1023.817	316.88	2584.34	Oxygen Delivery at Baseline (ml/min/m^2)
fate	455	2	.3868132	0	1	Mortal Status at 30 Days
followup	455	148	532.5824	1	720	Follow-up (hours)
temp0	455	122	100.4269	91.58	107	Baseline Temperature (deg. F)
temp1	420	106	99.84362	90.68	106.7	Temperature after 2 hours
temp2	402	108	99.54328	93.6	107.9	Temperature after 4 hours
temp3	418	113	99.38531	92.3	104.7	Temperature after 8 hours
temp4	421	111	99.23259	90.3	104.54	Temperature after 12 hours
temp5	422	113	99.01384	91.58	104.4	Temperature after 16 hours
temp6	432	108	99.10542	88.88	104.5	Temperature after 20 hours
temp7	413	105	99.19448	88.7	104.18	Temperature after 24 hours
temp8	407	105	99.19681	93.38	103.64	Temperature after 28 hours
temp9	401	102	99.14988	93.2	104.36	Temperature after 32 hours
temp10	399	101	99.16607	92.3	104	Temperature after 36 hours
temp11	402	98	99.12139	92.12	103.4	Temperature after 40 hours
temp12	406	97	99.09759	91.22	103.28	Temperature after 44 hours
temp13	403	104	99.47588	93.7	104.9	Temperature after 72 hours
temp14	316	87	99.29114	92.2	103.28	Temperature after 96 hours
temp15	382	93	99.42188	95	105.1	Temperature after 120 hours

```
. *****  Preliminary:  Display codes for race for later selection of race " other
. numlabel, add
. table race
```

Race	Freq.
0. White	293
1. Black	130
2. Other	32

```
. *
. *****  5.)  Single sample paired data, continuous:  Paired t-test
. *****  Consider race="other" ONLY
. *****  ttest var1==var2 if CONDITION
. ttest temp0==temp1 if race==2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   temp0 |      28    100.6436    .4145246    2.193458    99.79304    101.4941
   temp1 |      28    99.98143    .3400388    1.799316    99.28373    100.6791
---------+--------------------------------------------------------------------
    diff |      28    .6621432    .2233354     1.18178    .2038967     1.12039
------------------------------------------------------------------------------
     mean(diff) = mean(temp0 - temp1)                              t =   2.9648
 Ho: mean(diff) = 0                               degrees of freedom =       27

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.9969         Pr(|T| > |t|) = 0.0063          Pr(T > t) = 0.0031
```

INTERPRETATION:
In the sub-group of 32 who are classified as race='other', data on temp0 and temp1 were complete for n=28, or 87.5%. Among these n=28, the mean temperatures at baseline and two-hours were 100.6 and 99.9, respectively. The mean and standard deviation of the baseline to two-hour change in temperature were 0.67 and 0.22, respectively. A two-sided paired t-test of the null hypothesis that the mean change $\mu_d = 0$ yielded a paired t-statistic value of 2.96 and associated two sided p-value = 0.006. This is statistically significant. The null hypothesis is rejected. Conclude that these data provide statistically significant evidence that the change in temperature between the baseline and 2-hour occasions of measurement
$\mu_d \neq 0$.

```
. *
. ******    6.)  Single sample paired data, continuous:  90% Conf Interval of change
. ******    ttest var1==var2, level(##)
. *****     Still working with race="other" ONLY

. ttest temp0==temp1 if race==2, level(90)

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [90% Conf. Interval]
---------+--------------------------------------------------------------------
   temp0 |      28    100.6436    .4145246    2.193458    99.93752    101.3496
   temp1 |      28    99.98143    .3400388    1.799316    99.40224    100.5606
---------+--------------------------------------------------------------------
    diff |      28    .6621432    .2233354     1.18178    .2817385    1.042548
------------------------------------------------------------------------------
     mean(diff) = mean(temp0 - temp1)                              t =   2.9648
 Ho: mean(diff) = 0                               degrees of freedom =       27

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.9969         Pr(|T| > |t|) = 0.0063          Pr(T > t) = 0.0031
```

INTERPRETATION:
Based on this sample of n=28, with 90% confidence, the unknown mean change in temperature baseline → two hours among persons with race classification 'other' is estimated to be between 0.28 and 1.04.

```
. *
. ******   7.)  Full data set - Two independent samples, continuous:  2 Sample t test


. ******   sort groupvariable
. ******   ttest variable, by(groupvariable)

. sort treat
. ttest apache, by(treat)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
0. Place |     230    15.18696     .456478    6.922831    14.28752    16.08639
1. Ibupr |     224    15.47768    .4852049    7.261882    14.52151    16.43385
---------+--------------------------------------------------------------------
combined |     454     15.3304    .3325528    7.085794    14.67686    15.98393
---------+--------------------------------------------------------------------
    diff |           -.290722    .6657587               -1.599088    1.017644
------------------------------------------------------------------------------
    diff = mean(0. Place) - mean(1. Ibupr)                        t =  -0.4367
Ho: diff = 0                                     degrees of freedom =      452

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3313         Pr(|T| > |t|) = 0.6626          Pr(T > t) = 0.6687
```

INTERPRETATION:
In this cohort of n_0 = 230 treated with placebo and n_1 = 224 treated with ibuprofen, the mean APACHE scores were 15.19 (s_1 = 6.9) and 15.48 (s_2 = 7.3), respectively. A two sample t-test of the null hypothesis of equality of means yielded a t-statistic value of -0.44 and associated two sided p-value = 0.66. This is not statistically significant. The null hypothesis is not rejected. Conclude that these data do not provide statistically significant evidence of a treatment effect on APACHE score.

```
. *
. ******  8. Full Data Set - Two independent samples, continuous:  95% CI for difference
. ******  ttest variable, by(groupvariable) level(##)

. ttest apache, by(treat) level(95)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
0. Place |     230    15.18696     .456478    6.922831    14.28752    16.08639
1. Ibupr |     224    15.47768    .4852049    7.261882    14.52151    16.43385
---------+--------------------------------------------------------------------
combined |     454     15.3304    .3325528    7.085794    14.67686    15.98393
---------+--------------------------------------------------------------------
    diff |           -.290722    .6657587               -1.599088    1.017644
------------------------------------------------------------------------------
    diff = mean(0. Place) - mean(1. Ibupr)                        t =  -0.4367
Ho: diff = 0                                     degrees of freedom =      452

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3313         Pr(|T| > |t|) = 0.6626          Pr(T > t) = 0.6687
```

INTERPRETATION:
Based on this cohort of n_0 = 230 treated with placebo and n_1 = 224 treated with ibuprofen, respectively, with 95% confidence, the unknown difference in APACHE score (placebo minus ibuprofen) is estimated to be between -1.60 and +1.02.